

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
21 March 2002 (21.03.2002)

PCT

(10) International Publication Number  
**WO 02/22869 A2**

(51) International Patent Classification: **C12Q 1/68**

(21) International Application Number: **PCT/GB01/04108**

(22) International Filing Date:  
13 September 2001 (13.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0022458.4 13 September 2000 (13.09.2000) GB  
60/283,771 13 April 2001 (13.04.2001) US  
60/285,501 20 April 2001 (20.04.2001) US

(71) Applicant (for all designated States except US): **MEDICAL RESEARCH COUNCIL** [GB/GB]; 20 Park Crescent, London WN1 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HOLLIGER**, Philipp [CH/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **GHADESSY**, Farid [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB).

(74) Agents: **MASCHIO**, Antonio et al.; D Young & Co, 21 New Fetter Lane, London EC4A 1AD (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declaration under Rule 4.17:**

— of inventorship (Rule 4.17(iv)) for US only

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **DIRECTED EVOLUTION METHOD**

(57) Abstract: We describe a method of selecting an enzyme having replicase activity, the method comprising the steps of: (a) providing a pool of nucleic acids comprising members each encoding a replicase or a variant of the replicase; (b) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool together with the replicase or variant encoded by the nucleic acid member; (c) allowing nucleic acid replication to occur; and (d) detecting amplification of the nucleic acid member by the replicase. Methods for selecting agents capable of modulating replicase activity, and for selecting interacting polypeptides are also disclosed.

WO 02/22869 A2

## DIRECTED EVOLUTION METHOD

FIELD OF THE INVENTION

The present invention relates to methods for use in *in vitro* evolution of molecular libraries. In particular, the present invention relates to methods of selecting nucleic acids  
5 encoding gene products in which the nucleic acid and the activity of the encoded gene product are linked by compartmentalisation.

BACKGROUND TO THE INVENTION

Evolution requires the generation of genetic diversity (diversity in nucleic acid) followed by the selection of those nucleic acids which encode beneficial characteristics.  
10 Because the activity of the nucleic acids and their encoded gene product are physically linked in biological organisms (the nucleic acids encoding the molecular blueprint of the cells in which they are confined), alterations in the genotype resulting in an adaptive change(s) of phenotype produce benefits for the organism resulting in increased survival and offspring. Multiple rounds of mutation and selection can thus result in the progressive enrichment of  
15 organisms (and the encoding genotype) with increasing adaptation to a given selection condition. Systems for rapid evolution of nucleic acids or proteins *in vitro* must mimic this process at the molecular level in that the nucleic acid and the activity of the encoded gene product must be linked and the activity of the gene product must be selectable.

20 *In vitro* selection technologies are a rapidly expanding field and often prove more powerful than rational design to obtain biopolymers with desired properties. In the past decade selection experiments, using e.g. phage display or SELEX technologies have yielded many novel polynucleotide and polypeptide ligands. Selection for catalysis has proved harder. Strategies have included binding of transition state analogues, covalent linkage to suicide  
25 inhibitors, proximity coupling and covalent product linkage. Although these approaches focus only on a particular part of the enzymatic cycle, there have been some successes. Ultimately however it would be desirable to select directly for catalytic turnover. Indeed, simple screening for catalytic turnover of fairly small mutant libraries has been rather more

successful than the various selection approaches and has yielded some catalysts with greatly improved catalytic rates.

While polymerases are a prerequisite for technologies that define molecular biology, i.e. site-directed mutagenesis, cDNA cloning and in particular Sanger sequencing and PCR, they often suffer from serious shortcomings due to the fact that they are made to perform tasks for which nature has not optimized them. Few attempts appear to have been made to improve the properties of polymerases available from nature and to tailor them for specific applications by protein engineering. Technical advances have been largely peripheral, and include the use of polymerases from a wider range of organisms, buffer and additive systems as well as enzyme blends.

Attempts to improve the properties of polymerases have traditionally relied on protein engineering. For example, variants of *Taq* polymerase (for example, Stoffel fragment and KlenTaq) have been generated by full or partial deletion of its 5'-3' exonuclease domain and show improved thermostability and fidelity although at the cost of reduced processivity (Barnes 1992, *Gene* 112, 29-35, Lawyer et al, 1993, *PCR Methods and Applications* 2, 275). In addition, the availability of high-resolution structures for proteins has allowed the rational design of mutants with improved properties (for example, *Taq* mutants with improved properties of dideoxynucleotide incorporation for cycle sequencing, Li et al., 1999, *Proc. Natl. Acad. Sci USA* 96, 9491). *In vivo* genetic approaches have also been used for protein design, for example by complementation of a *polA*<sup>-</sup> strain to select for active polymerases from repertoires of mutant polymerases (Suzuki et al., 1996 *Proc. Natl. Acad. Sci USA* 93, 9670). However, the genetic complementation approach is limited in the properties that can be selected for.

Recent advances in molecular biology have allowed some molecules to be co-selected *in vitro* according to their properties along with the nucleic acids that encode them. The selected nucleic acids can subsequently be cloned for further analysis or use, or subjected to additional rounds of mutation and selection. Common to these methods is the establishment of large libraries of nucleic acids. Molecules having the desired characteristics (activity) can be

isolated through selection regimes that select for the desired activity of the encoded gene product, such as a desired biochemical or biological activity, for example binding activity.

WO99/02671 describes a method for isolating one or more genetic elements encoding a gene product having a desired activity. Genetic elements are first compartmentalised into microcapsules, and then transcribed and/or translated to produce their respective gene products (RNA or protein) within the microcapsules. Alternatively, the genetic elements are contained within a host cell in which transcription and/or translation (expression) of the gene product takes place and the host cells are first compartmentalised into microcapsules. Genetic elements which produce gene product having desired activity are subsequently sorted. The method described in WO99/02671 relies on the gene product catalytically modifying the microcapsule or the genetic element (or both), so that enrichment of the modified entity or entities enables selection of the desired activity.

#### SUMMARY OF THE INVENTION

According to a first aspect of the present invention, we provide a method of selecting a nucleic acid-processing (NAP) enzyme, the method comprising the steps of: (a) providing a pool of nucleic acids comprising members encoding a NAP enzyme or a variant of the NAP enzyme; (b) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool together with the NAP enzyme or variant encoded by the nucleic acid member; (c) allowing nucleic acid processing to occur; and (d) detecting processing of the nucleic acid member by the NAP enzyme.

There is provided, according to a second aspect of the present invention, a method of selecting an agent capable of modifying the activity of a NAP enzyme, the method comprising the steps of: (a) providing a NAP enzyme; (b) providing a pool of nucleic acids comprising members encoding one or more candidate agents; (c) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool, the agent encoded by the nucleic acid member, and the NAP enzyme; and (d) detecting processing of the nucleic acid member by the NAP enzyme.



Preferably, the agent is a promoter of NAP enzyme activity. The agent may be an enzyme, preferably a kinase or a phosphorylase, which is capable of acting on the NAP enzyme to modify its activity. The agent may be a chaperone involved in the folding or assembly of the NAP enzyme or required for the maintenance of replicase function (e.g. telomerase, HSP 90). Alternatively, the agent may be a polypeptide or polynucleotide involved in a metabolic pathway, the pathway having as an end product a substrate which is involved in a replication reaction. The agent may moreover be any enzyme which is capable of catalysing a reaction that modifies an inhibiting agent (natural or unnatural) of the NAP enzyme in such a way as to reduce or abolish its inhibiting activity. Finally the agent may promote NAP activity in a non-catalytic way, e.g. by association with the NAP enzyme or its substrate etc. (e.g. processivity factors in the case of DNA polymerases, e.g. T7 DNA polymerase & thioredoxin).

We provide, according to a third aspect of the present invention, a method of selecting a pair of polypeptides capable of stable interaction, the method comprising: (a) providing a first nucleic acid and a second nucleic acid, the first nucleic acid encoding a first fusion protein comprising a first subdomain of a NAP enzyme fused to a first polypeptide, the second nucleic acid encoding a second fusion protein comprising a second subdomain of a NAP enzyme fused to a second polypeptide; in which stable interaction of the first and second NAP enzyme subdomains generates NAP enzyme activity, and in which at least one of the first and second nucleic acids is provided in the form of a pool of nucleic acids encoding variants of the respective first and/or second polypeptide(s); (b) subdividing the pool or pools of nucleic acids into compartments, such that each compartment comprises a first nucleic acid and a second nucleic acid together with respective fusion proteins encoded by the first and second nucleic acids; (c) allowing the first polypeptide to bind to the second polypeptide, such that binding of the first and second polypeptides leads to stable interaction of the NAP enzyme subdomains to generate NAP enzyme activity; and (d) detecting processing of at least one of the first and second nucleic acids by the NAP enzyme.

Moreover, the NAP enzyme domains referred to in (a) above may be replaced with domains of a polypeptide capable of modifying the activity of NAP enzymes, as discussed in

the second aspect of the present invention, and NAP enzyme activity used to select such modifying polypeptides having desired properties.

Preferably, each of the first and second nucleic acids is provided from a pool of nucleic acids.

5 Preferably, the first and second nucleic acids are linked either covalently (e.g. as part of the same template molecule) or non-covalently (e.g. by tethering onto beads etc.).

NAP enzymes may for example be polypeptide or ribonucleic acid enzyme molecules. In a highly preferred embodiment, the NAP enzyme according to the invention is a replicase enzyme, i.e. an enzyme, which is capable of amplifying nucleic acid from a template, such as  
10 for example a polymerase enzyme (or ligase). The invention is described herein below with specific reference to replicases; however, it will be understood by those skilled in the art that the invention is equally applicable to other NAP enzymes, such as telomerases and helicases, as further set out below, which process nucleic acids in ways not limited to amplification but which are nevertheless selectable by detecting nucleic acid amplification, i.e. which promote  
15 replication indirectly.

In a preferred embodiment of the invention, amplification of the nucleic acid results from more than one round of nucleic acid replication. Preferably, the amplification of the nucleic acid is an exponential amplification.

The amplification reaction is preferably selected from the following: a polymerase  
20 chain reaction (PCR), a reverse transcriptase-polymerase chain reaction (RT-PCR), a nested PCR, a ligase chain reaction (LCR), a transcription based amplification system (TAS), a self-sustaining sequence replication (3SR), NASBA, a transcription-mediated amplification reaction (TMA), and a strand-displacement amplification (SDA).

In a highly preferred embodiment, the post-amplification copy number of the nucleic  
25 acid member is substantially proportional to the activity of the replicase, the activity of a requisite agent, binding affinity of the first and second polypeptides.

Nucleic acid replication may be detected by assaying the copy number of the nucleic acid member. Alternatively, or in addition, nucleic acid replication may be detected by determining the activity of a polypeptide encoded by the nucleic acid member.

5 In a highly preferred embodiment, the conditions in the compartment are adjusted to select for a replicase or agent active under such conditions, or a pair of polypeptides capable of stable interaction under such conditions.

The replicase preferably has polymerase, reverse transcriptase or ligase activity.

10 The polypeptide may be provided from the nucleic acid by *in vitro* transcription and translation. Alternatively, the polypeptide may be provided from the nucleic acid *in vivo* in an expression host.

In a preferred embodiment, the compartments consist of the encapsulated aqueous component of a water-in-oil emulsion. The water-in-oil emulsion is preferably produced by emulsifying an aqueous phase with an oil phase in the presence of a surfactant comprising 4.5% v/v Span 80, 0.4% v/v Tween 80 and 0.1% v/v Triton X100, or a surfactant comprising 15 Span 80, Tween 80 and Triton X100 in substantially the same proportions. Preferably, the water:oil phase ratio is 1:2, which leads to adequate droplet size. Such emulsions have a higher thermal stability than more oil-rich emulsions.

20 As a fourth aspect of the present invention, there is provided a replicase enzyme identified by a method according to any preceding claim. Preferably, the replicase enzyme has a greater thermostability than a corresponding unselected enzyme. More preferably, the replicase enzyme is a *Taq* polymerase having more than 10 times increased half-life at 97.5 °C when compared to wild type *Taq* polymerase.

25 The replicase enzyme may have a greater tolerance to heparin than a corresponding unselected enzyme. Preferably, the replicase enzyme is a *Taq* polymerase active at a concentration of 0.083 units/ $\mu$ l or more of heparin

The replicase enzyme may be capable of extending a primer having a 3' mismatch. Preferably, the 3' mismatch is a 3' purine-purine mismatch or a 3' pyrimidine-pyrimidine mismatch. More preferably, the 3' mismatch is an A-G mismatch or the 3' mismatch is a C-C mismatch.

5 We provide, according to a fifth aspect of the present invention, a *Taq* polymerase mutant comprising the mutations (amino acid substitutions): F73S, R205K, K219E, M236T, E434D and A608V.

The present invention, in a sixth aspect, provides a *Taq* polymerase mutant comprising the mutations (amino acid substitutions): K225E, E388V, K540R, D578G, N583S and  
10 M747R.

The present invention, in a seventh aspect, provides a *Taq* polymerase mutant comprising the mutations (amino acid substitutions): G84A, D144G, K314R, E520G, A608V, E742G.

The present invention, in a eighth aspect, provides a *Taq* polymerase mutant comprising the mutations (amino acid substitutions): D58G, R74P, A109T, L245R, R343G, G370D, E520G,  
15 N583S, E694K, A743P.

In a ninth aspect of the present invention, there is provided a water-in-oil emulsion obtainable by emulsifying an aqueous phase with an oil phase in the presence of a surfactant comprising 4.5% v/v Span 80, 0.4% v/v Tween 80 and 0.1% v/v Triton X100, or a surfactant comprising Span 80, Tween 80 and Triton X100 in substantially the same proportions.  
20 Preferably, the water:oil phase ratio is 1:2. This ratio appears to permit diffusion of dNTPs (and presumably other small molecules) between compartments at higher temperatures, which is beneficial for some applications but not for others. Diffusion can be controlled by increasing water:oil phase ratio to 1:4.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a diagram showing an embodiment of a method according to the present invention as applied to selection of a self-evolving polymerase, in which gene copy number is linked to enzymatic turnover.

5        Figure 1B is a diagram showing a general scheme of compartmentalised self-replication (CSR): 1) A repertoire of diversified polymerase genes is cloned and expressed in *E.coli*. Spheres represent active polymerase molecules. 2) Bacterial cells containing the polymerase and encoding gene are suspended in reaction buffer containing flanking primers and nucleotide triphosphates (dNTPs) and segregated into aqueous compartments. 3) The  
10        polymerase enzyme and encoding gene are released from the cell allowing self-replication to proceed. Poorly active polymerases (white hexagon) fail to replicate their encoding gene. 4) The "offspring" polymerase genes are released, rediversified and recloned for another cycle of CSR.

15        Figure 2 is a diagram showing aqueous compartments of the heat-stable emulsion containing *E.coli* cells expressing green fluorescent protein (GFP) prior to (A, B), and after thermocycling (C), as imaged by light microscopy. (A, B) represent the same frame. (A) is imaged at 535 nm for GFP fluorescence and (B) in visible light to visualize bacterial cells within compartments. Smudging of the fluorescent bacteria in (A) is due to Brownian motion during exposure. Average compartment dimensions as determined by laser diffraction are  
20        given below.

Figure 3A is a diagram showing crossover between emulsion compartments. Two standard PCR reactions, differing in template size (PCR1 (0.9kb), PCR2 (0.3 kb)) and presence of *Taq* (PCR1: + *Taq*, PCR 2: no enzyme), are amplified individually or combined. When combined in solution, both templates are amplified. When emulsified separately, prior  
25        to mixing, only PCR1 is amplified. M:  $\phi$ X174 *Hae*III marker

Figure 3B is a diagram showing crossover between emulsion compartments. Bacterial cells expressing wild-type *Taq* polymerase (2.7kb) or the *Taq* polymerase Stoffel fragment

(poorly active under the buffer conditions) (1.8kb) are mixed 1:1 prior to emulsification. In solution, the shorter Stoffel fragment is amplified preferentially. In emulsion, there is predominantly amplification of the wt *Taq* gene and only weak amplification of the Stoffel fragment (arrow). M:  $\lambda$ HindIII marker

5        Figure 4 is a diagram showing details of an embodiment of a method according to the present invention as applied to selection of a self-evolving polymerase.

Figure 5 is a diagram showing details of an embodiment of a method according to the present invention to select for incorporation of novel or unusual substrates.

10        Figure 6 is a diagram showing selection of RNA having (intermolecular) catalytic activity using the methods of our invention.

Figure 7 is a diagram showing a model of a *Taq*-DNA complex.

Figure 8: A: General scheme of a cooperative CSR reaction.

15        Nucleoside diphosphate kinase (ndk) is expressed from a plasmid and converts deoxynucleoside diphosphates which are not substrates for *Taq* polymerase into deoxynucleoside triphosphates which are. As soon as ndk has produced sufficient amounts of substrate, *Taq* can replicate the ndk gene.

20        B: Bacterial cells expressing wild-type ndk (0.8kb) or an inactive truncated fragment (0.5kb) are mixed 1:1 prior to emulsification. In solution, the shorter truncated fragment is amplified preferentially. In emulsion, there is predominantly amplification of the wt ndk gene and only weak amplification of the truncated fragment (arrow) indicating that in emulsion only active ndk genes producing substrate are amplified. M:  
25        *Hae*III  $\phi$ X174 marker

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art.

5 Such techniques are explained in the literature. See, e.g., J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; J. M. Polak and James O'D. McGee, 1990, *In Situ Hybridization: Principles and Practice*; Oxford University Press; M. J. Gait

10 (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology*, Academic Press. Each of these general texts are herein incorporated by reference.

## COMPARTMENTALISED SELF REPLICATION

15

Our invention describes a novel selection technology, which we call CSR (compartmentalised self-replication). It has the potential to be expanded into a generic selection system for catalysis as well as macromolecular interactions.

20

In its simplest form CSR involves the segregation of genes coding for and directing the production of DNA polymerases within discrete, spatially separated, aqueous compartments of a novel heat-stable water-in-oil emulsion. Provided with nucleotide triphosphates and appropriate flanking primers, polymerases replicate only their own genes. Consequently, only genes encoding active polymerases are replicated, while inactive variants

25 that cannot copy their genes disappear from the gene pool. By analogy to biological systems, among differentially adapted variants, the most active (the fittest) produce the most "offspring", hence directly correlating post-selection copy number with enzymatic turn-over.

CSR is not limited to polymerases but can be applied to a wide variety of enzymatic

30 transformations, built around the "replicase engine". For example, an enzyme "feeding" a

polymerase which in turn replicates its gene may be selected. More complicated coupled cooperative reaction schemes can be envisioned in which several enzymes either produce replicase substrates or consume replicase inhibitors.

5 Polymerases occupy a central role in genome maintenance, transmission and expression of genetic information. Polymerases are also at the heart of modern biology, enabling core technologies such as mutagenesis, cDNA libraries, sequencing and the polymerase chain reaction (PCR). However, commonly used polymerases frequently suffer from serious shortcomings as they are used to perform tasks for which nature had not  
10 optimized them. Indeed, most advances have been peripheral, including the use of polymerases from different organisms, improved buffer and additive systems as well as enzyme blends. CSR is a novel selection system ideally suited for the isolation of "designer" polymerases for specific applications. Many features of polymerase function are open to "improvement" (e.g. processivity, substrate selection etc.). Furthermore, CSR is a tool to  
15 study polymerase function, e.g. to probe immutable regions, study components of the replisome etc. Moreover, CSR may be used for shotgun functional cloning of polymerases, straight from diverse, uncultured microbial populations.

CSR represents a novel principle of repertoire selection of polypeptides. Previous  
20 approaches have featured various "display" methods in which phenotype and genotype (polypeptide and encoding gene) are linked as part of a "genetic package" containing the encoding gene and displaying the polypeptide on the "outside". Selection occurs via a step of affinity purification after which surviving clones are grown (amplified) in cells for further rounds of selection (with resulting biases in growth distorting selections). Further distortions  
25 result from differences in the display efficiencies between different polypeptides.

In another set of methods both polypeptide and encoding gene(s) are "packaged" within a cell. Selection occurs *in vivo* through the polypeptide modifying the cell in such a way that it acquires a novel phenotype, e.g. growth in presence of an antibiotic. As the selection pressure  
30 is applied on whole cells, such approaches tend to be prone to the generation of false positives. Furthermore, *in vivo* complementation strategies are limited in that selection



conditions, and hence selectable phenotypes, cannot be freely chosen and are further constrained by limits of host viability.

5 In CSR, there is no direct physical linkage (covalent or non-covalent) between polypeptide and encoding gene. More copies of successful genes are "grown" directly and *in vitro* as part of the selection process.

10 CSR is applicable to a broad spectrum of DNA and RNA polymerases, indeed to all polypeptides (or polynucleotides) involved in replication or gene expression. CSR can also be applied to DNA and RNA ligases assembling their genes from oligonucleotide fragments.

CSR is the only selection system in which the turn-over rate of an enzyme is directly linked to the post-selection copy-number of its encoding gene.

15 There is great interest in polynucleotide polymers with altered bases, altered sugars or even backbone chemistries. However, solid-phase synthesis can usually only provide relatively short polymers and naturally occurring polymerases unsurprisingly incorporate most analogues poorly. CSR is ideally suited for the selection of polymerases more tolerant of unnatural substrates in order to prepare polynucleotide polymers with novel properties for chemistry, biology and nanotechnology (e.g. DNA wires).

20 Finally, the heat-stable emulsion developed for CSR has applications on its own. With  $> 10^9$  microcompartments/ml, emulsion PCR (ePCR) offers the possibility of parallel PCR multiplexing on a unprecedented scale with potential applications from gene linkage analysis to genomic repertoire construction directly from single cells. It may also have applications for large-scale diagnostic PCR applications like "Digital PCR" (Vogelstein & Kinzler (1999), *PNAS*, 96, 9236-9241). Compartmentalizing individual reactions can also even out competition among different gene segments that are amplified in either multiplex or random primed PCR and leads to a less biased distribution of amplification products. ePCR may thus provide an alternative to whole genome DOP-PCR (and related methodologies) or indeed be used to make DOP-PCR (and related methodologies) more effective.

The selection system according to our invention is based on self-replication in a compartmentalised system. Our invention relies on the fact that active replicases are able to replicate nucleic acids (in particular their coding sequences), while inactive replicases cannot. Thus, in the methods of our invention, we provide a compartmentalised system where a replicase in a compartment is substantially unable to act on any template other than the templates within that compartment; in particular, it cannot act to replicate a template within any other compartment. In highly preferred embodiments, the template nucleic acid within the compartment encodes the replicase. Thus, the replicase cannot replicate anything other than its coding sequence; the replicase is therefore "linked" to its coding sequence. As a result, in highly preferred embodiments of our invention, the final concentration of the coding sequence (i.e. copy number) is dependent on the activity of the enzyme encoded by it.

Our selection system as applied to selection of replicases has the advantage in that it links catalytic turnover ( $k_{cat}/K_m$ ) to the post-selection copy-number of the gene encoding the catalyst. Thus, compartmentalisation offers the possibility of linking genotype and phenotype of a replicase enzyme, as described in further detail below, by a coupled enzymatic reaction involving the replication of the gene or genes of the enzyme(s) as one of its steps.

The methods of our invention preferably make use of nucleic acid libraries, the nature and construction of which will be explained in greater detail below. The nucleic acid library comprises a pool of different nucleic acids, members of that encode variants of a particular entity (the entity to be selected). Thus, for example, as used to select for replicases, the methods of our invention employ a nucleic acid library or pool having members, which encode the replicase or variants of the replicase. Each of the entities encoded by the various members of the library will have different properties, e.g., varying tolerance to heat or to the presence of inhibitory small molecules, or tolerance for base pair mismatches (as explained in further detail below). The population of nucleic acid variants therefore provides a starting material for selection, and is in many ways analogous to variation in a natural population of organisms caused by mutation.

According to our invention, the different members of the nucleic acid library or pool are sorted or compartmentalised into many compartments or microcapsules. In preferred

embodiments, each compartment contains substantially one nucleic acid member of the pool (in one or several copies). In addition, the compartment also comprises the polypeptide or polynucleotide (in one or preferably several copies) encoded by that nucleic acid member (whether it is a replicase, an agent, a polypeptide, etc as discussed below). The nature of these  
5 compartments is such that minimal or substantially no interchange of macromolecules (such as nucleic acids and polypeptides) occurs between different compartments. As explained in further detail below, highly preferred embodiments of our invention make use of aqueous compartments within water-in-oil emulsions. As explained above, any replicase activity present in the compartment (whether exhibited by the replicase, modified by an agent, or  
10 exhibited by the polypeptide acting in conjunction with another polypeptide) can only act on the template within the compartment.

The conditions within the compartments may be varied in order to select for polypeptides active under these conditions. For example, where replicases are selected, the compartments may have an increased temperature to select for replicases with higher thermal  
15 stability. Furthermore, using the selection methods described here on fusion proteins comprising thermostable replicase and a protein of interest will allow the selection of thermally stable proteins.

A method for the incorporation of thermal stability into otherwise labile proteins of  
20 commercial importance is desirable with regards to their large-scale production and distribution. A reporter system has been described to improve protein folding by expressing proteins as fusions with green fluorescent protein (GFP) (Waldo et al (1999), *Nat Biotechnol* 17: 691-695). The function of the latter is related to the productive folding of the fused protein influencing folding and/or functionality of the GFP, enabling the directed evolution of  
25 variants with improved folding and expression. According to this aspect of our invention, proteins are fused to a thermostable replicase (or an agent promoting replicase activity) and selecting for active fusions in emulsion as a method for evolving proteins with increased thermostability and/or solubility. Unstable variants of the fusion partner are expected to aggregate and precipitate prior to or during thermal cycling, thus compromising replicase  
30 activity within respective compartments. Viable fusions will allow for self-amplification in emulsion, with the turn-over rate being linked to the stability of the fusion partner.

In a related approach, novel or increased chaperonin activity may be evolved by coexpression of a library of chaperones together with a polymerase-polypeptide fusion protein, in which the protein moiety misfolds (under the selection conditions). Replication of the gene(s) encoding the chaperonin can only proceed after chaperonin activity has rescued  
5 polymerase activity in the polymerase-polypeptide fusion protein.

Thermostability of an enzyme may be measured by conventional means as known in the art. For example, the catalytic activity of the native enzyme may be assayed at a certain temperature as a benchmark. Enzyme assays are well known in the art, and standard assays have been established over the years. For example, incorporation of nucleotides by a  
10 polymerase is measured, by for example, use of radiolabeled dNTPs such as dATP and filter binding assays as known in the art. The enzyme whose thermostability is to be assayed is preincubated at an elevated temperature and then its activity retained. (for example, polymerase activity in the case of polymerases) is measured at a lower, optimum temperature and compared to the benchmark. In the case of *Taq* polymerase, the elevated temperature is  
15 97.5°C; the optimum temperature is 72°C. Thermostability may be expressed in the form of half-life at the elevated temperature (i.e. time of incubation at higher temperature over which polymerase loses 50% of its activity). For example, the thermostable replicases, fusion proteins or agents selected by our invention may have a half-life that is 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X or more than the native enzyme. Most preferably, the thermostable  
20 replicases etc have a half-life that is 11x or more when compared this way. Preferably, selected polymerases are preincubated at 95 °C or more, 97.5 °C or more, 100 °C or more, 105 °C or more, or 110 °C or more. Thus, in a highly preferred embodiment of our invention, we provide polymerases with increased thermostability which display a half life at 97.5 °C that is 11X or more than the corresponding wild type (native) enzyme.

25 Resistance to an inhibitory agent, such as heparin in the case of polymerases, may also be assayed and measured as above. Resistance to inhibition may be expressed in terms of the concentration of the inhibitory factor. For example, in preferred embodiments of the invention, we provide heparin resistant polymerases that are active in up to a concentration of heparin between 0.083units/μl to 0.33 units/μl. For comparison, our assays indicate that the

concentration of heparin which inhibits native (wild-type) *Taq* polymerase is in the region of between 0.0005 to 0.0026 units/ $\mu$ l.

Resistance is conveniently expressed in terms of the inhibitor concentration, which is found to inhibit the activity of the selected replicase, fusion protein or agent, compared to the concentration, which is found to inhibit the native enzyme. Thus, the resistant replicases, fusion proteins, or agents selected by our invention may have 10X, 20X, 30X, 40X, 50X, 60X, 70X, 80X, 90X, 100X, 110X, 120X, 130X, 140X, 150X, 160X, 170X, 180X, 190X, 200X, or more resistance compared to the native enzyme. Most preferably, the resistant replicases etc have 130x or more fold increased resistance when compared this way. The selected replicases etc preferably have 50% or more, 60% or more, 70% or more, 80% or more, 90% or more, or even 100% activity at the concentration of the inhibitory factor. Furthermore, the compartments may contain amounts of an inhibitory agent such as heparin to select for replicases having activity under such conditions.

As explained below, the methods of our invention may be used to select for a pair of interacting polypeptides, and the conditions within the compartments may be altered to choose polypeptides capable of acting under these conditions (for example, high salt, or elevated temperature, etc.). The methods of our invention may also be used to select for the folding, stability and/or solubility of a fused polypeptide acting under these conditions (for example, high salt, or elevated temperature, chaotropic agents etc.).

The method of selection of our present invention may be used to select for various replicative activities, for example, for polymerase activity. Here, the replicase is a polymerase, and the catalytic reaction is the replication by the polymerase of its own gene. Thus, defective polymerases or polymerases which are inactive under the conditions under which the reaction is carried out (the selection conditions) are unable to amplify their own genes. Similarly, polymerases which are less active will replicate their coding sequences within their compartments more slowly. Accordingly, these genes will be under-represented, or even disappear from the gene pool.

Active polymerases, on the other hand, are able to replicate their own genes, and the resulting copy number of these genes will be increased. In a preferred embodiment of the invention, the copy number of a gene within the pool will bear a direct relation to the activity of the encoded polypeptide under the conditions under which the reaction is carried out. In this preferred embodiment, the most active polymerase will be most represented in the final pool (i.e., its copy number within the pool will be highest). As will be appreciated, this enables easy cloning of active polymerases over inactive ones. The method of our invention therefore is able to directly link the turnover rate of the enzyme to the resulting copy-number of the gene encoding it.

As an example, the method may be applied to the isolation of active polymerases (DNA-, RNA-polymerases and reverse transcriptases) from thermophilic organisms. Briefly a thermostable polymerase is expressed intracellularly in bacterial cells and these are compartmentalised (e.g. in a water-oil emulsion) in appropriate buffer together with appropriate amounts of the four dNTPs and oligonucleotides priming at either end of the polymerase gene or on plasmid sequences flanking the polymerase gene. The polymerase and its gene are released from the cells by a temperature step that lyses the cells and destroys enzymatic activities associated with the host cell. Polymerases from mesophilic organisms (or less thermostable polymerases) may be expressed in an analogous way except cell lysis should either proceed at ambient temperature (e.g. by expression of a lytic protein (e.g. derived from lytic bacteriophages, by detergent mediated lysis (e.g. Bugbuster™, commercially available) or lysis may proceed at elevated temperature in the presence of a polymerase stabilizing agent (e.g. high concentrations of proline (see example 27) in the case of *Klenow* or trehalose in the case of RT). In such cases background polymerase activity of the host strain may interfere with selections and it may be preferable to make use of mutant strains (e.g. *polA*<sup>-</sup>).

Alternatively, polymerase genes (either as plasmids or linear fragments) may be compartmentalised as above and the polymerase expressed *in situ* within the compartments using *in vitro* transcription translation (*ivt*), followed by a temperature step to destroy enzymatic activities associated with the *in vitro* translation extract. Polymerases from mesophilic organisms (or less thermostable polymerases) may be expressed *in situ* in an analogous way except in order to avoid enzymatic activities associated with the *in vitro*

translation extract it may be preferable to use a translation extract reconstituted from defined purified components like the PURE system (Shimizu *et al* (2001) *Nat. Biotech.*, 19:751).

5 PCR thermocycling then leads to the amplification of the polymerase genes by the polypeptides they encode, i.e. only genes encoding active polymerases, or polymerases active under the chosen conditions will be amplified. Furthermore, the copy number of a polymerase gene X after self-amplification will be directly proportional to the catalytic activity of the polymerase X it encodes. (see Figures 1A and 1B).

10 By varying the selection conditions within the compartment, polymerases or other replicases with desired properties may be selected using the methods of our invention. Thus, by exposing repertoires of polymerase genes (diversified through targeted or random mutation) to self-amplification and by altering the conditions under which self-amplification can occur, the system can be used for the isolation and engineering of polymerases with altered, enhanced or novel properties. Such enhanced properties may include increased 15 thermostability, increased processivity, increased accuracy (better proofreading), increased incorporation of unfavorable substrates (e.g., ribonucleotides, dye-modified, general bases such as 5-nitroindole, or other unusual substrates such as pyrene nucleotides (Matray & Kool (1999), *Nature* 399, 704-708) (Fig. 3) or resistance to inhibitors (e.g. Heparin in clinical samples). Novel properties may be the incorporation of unnatural substrates (e.g. 20 ribonucleotides), bypass reading of damaged sites (e.g. abasic sites (Paz-Elizur T. et al (1997) *Biochemistry* 36, 1766), thymidine-dimers (Wood R.D. (1999) *Nature* 399, 639), hydantoin-bases (Duarte V *et al* (1999) *Nucleic Acids Res.* 27, 496) and possibly even novel chemistries (e.g. novel backbones such as PNA (Nielsen PE. *Curr Opin Biotechnol.* 1999;10(1):71-5) or sulfone (Benner SA et al. *Pure Appl Chem.* 1998 Feb;70(2):263-6) or altered sugar chemistries 25 (A. Eschenmoser, *Science* 284, 2118-24 (1999)). It may also be used to isolate or evolve factors that enhance or modify polymerase function such as processivity factors (like thioredoxin in the case of T7 DNA polymerase (Doublié S. *et al* (1998) *Nature* 391, 251))

30 However, other enzymes besides replicases, such as telomerases, helicases etc may also be selected according to our invention. Thus, telomerase is expressed *in situ* (in

compartments) by for example *in vitro* translation together with Telomerase-RNA (either added or transcribed in situ as well; e.g. Bachand *et al.*, (2000) RNA 6:778-784).

Compartments also contain *Taq* Pol and dNTPs and telomere specific primers. At low temperature *Taq* is inactive but active telomerase will append telomeres to its own encoding gene (a linear DNA fragment with appropriate ends). After the telomerase reaction, thermocycling only amplifies active telomerase encoding genes. Diversity can be introduced in telomerase gene or RNA (or both) and could be targeted or random. As applied to selection of helicases, the selection method is essentially the same as described for telomerases, but helicase is used to unwind strands rather than heat denaturation

The methods of our invention may also be used to select for DNA repair enzymes or translesion polymerases such as *E.coli* Pol IV and Pol V. Here, damage is introduced into primers (targeted chemistry) or randomly by mutagen treatment (e.g. UV, mutagenic chemicals etc.). This allows for selection for enzymes able to repair primers required for replication or own gene sequence (information retrieval) or, resulting in improved "repairases" for gene therapy etc.

The methods of our invention may also be used in its various embodiments for selecting agents capable of directly or indirectly modulating replicase activity. In addition, the invention may be used to select for a pair of polypeptides capable of interacting, or for selection of catalytic nucleic acids such as catalytic RNA (ribozymes). These and other embodiments will be explained in further detail below.

#### NUCLEIC ACID PROCESSING ENZYMES

As referred to herein, a nucleic acid processing enzyme is any enzyme, which may be a protein enzyme or a nucleic acid enzyme, which is capable of modifying, extending (such as by at least one nucleotide), amplifying or otherwise influencing nucleic acids such as to render the nucleic acid selectable by amplification in accordance with the present invention. Such enzymes therefore possess an activity which results in, for example, amplification, stabilisation, destabilisation, hybridisation or denaturation, replication, protection or



deprotection of nucleic acids, or any other activity on the basis of which a nucleic acid can be selected by amplification. Examples include helicases, telomerases, ligases, recombinases, integrases and replicases. Replicases are preferred.

#### REPLICASE/REPLICATION

5 As used here, the term "replication" refers to the template-dependent copying of a nucleic acid sequence. Nucleic acids are discussed and exemplified below. In general, the product of the replication is another nucleic acid, whether of the same species, or of a different species. Thus, included are the replication of DNA to produce DNA, replication of DNA to produce RNA, replication of RNA to produce DNA and replication of RNA to  
10 produce RNA. "Replication" is therefore intended to encompass processes such as DNA replication, polymerisation, ligation of oligonucleotides or polynucleotides (e.g. tri-nucleotide (triplet) 5' triphosphates) to form longer sequences, transcription, reverse transcription, etc.

The term "replicase" is intended to mean an enzyme having catalytic activity, which is capable of joining nucleotide, building blocks together to form nucleic acid sequences. Such  
15 nucleotide building blocks include, but are not limited to, nucleosides, nucleoside triphosphates, deoxynucleosides, deoxynucleoside triphosphates, nucleotides (comprising a nitrogen-containing base such as adenine, guanine, cytosine, uracil, thymine, etc., a 5-carbon sugar and one or more phosphate groups), nucleotide triphosphates, deoxynucleotides such as deoxyadenosine, deoxythymidine, deoxycytidine, deoxyuridine, deoxyguanine,  
20 deoxynucleotides triphosphates (dNTPs), and synthetic or artificial analogues of these. Building blocks also include oligomers or polymers of any of the above, for example, trinucleotides (triplets), oligonucleotides and polynucleotides.

Thus, a replicase may extend a pre-existing nucleic acid sequence (primer) by incorporating nucleotides or deoxynucleotides. Such an activity is known in the art as  
25 "polymerisation", and the enzymes, which carry this out, are known as "polymerases". An example of such a polymerase replicase is DNA polymerase, which is capable of replicating DNA. The primer may be the same chemically, or different from, the extended sequence (for example, mammalian DNA polymerase is known to extend a DNA sequence from an RNA

primer). The term replicase also includes those enzymes which join together nucleic acid sequences, whether polymers or oligomers to form longer nucleic acid sequences. Such an activity is exhibited by the ligases, which ligate pieces of DNA or RNA.

The replicase may consist entirely of replicase sequence, or it may comprise a replicase sequence linked to a heterologous polypeptide or other molecule such as an agent by chemical means or in the form of a fusion protein or be assembled from two or more constituent parts.

Preferably, the replicase according to the invention is a DNA polymerase, RNA polymerase, reverse transcriptase, DNA ligase, or RNA ligase.

10 Preferably, the replicase is a thermostable replicase. A "thermostable" replicase as used here is a replicase, which demonstrates significant resistance to thermal denaturation at elevated temperatures, typically above body temperature (37°C). Preferably, such a temperature is in the range 42°C to 160°C, more preferably, between 60 to 100°C, most preferably, above 90°C. Compared to a non-thermostable replicase, the thermostable replicase  
15 displays a significantly increased half-life (time of incubation at elevated temperature that results in 50% loss of activity). Preferably, the thermostable replicase retains 30% or more of its activity after incubation at the elevated temperature, more preferably, 40%, 50%, 60%, 70% or 80% or more of its activity. Yet more preferably, the replicase retains 80% activity. Most preferably, the activity retained is 90%, 95% or more, even 100%. None-thermostable  
20 replicases would exhibit little or no retention of activity after similar incubations at the elevated temperature.

#### POLYMERASE

An example of a replicase is DNA polymerase. DNA polymerase enzymes are naturally occurring intracellular enzymes, and are used by a cell to replicate a nucleic acid  
25 strand using a template molecule to manufacture a complementary nucleic acid strand. Enzymes having DNA polymerase activity catalyze the formation of a bond between the 3' hydroxyl group at the growing end of a nucleic acid primer and the 5' phosphate group of a

nucleotide triphosphate. These nucleotide triphosphates are usually selected from deoxyadenosine triphosphate (A), deoxythymidine triphosphate (T), deoxycytidine triphosphate (C) and deoxyguanosine triphosphate (G). However, DNA polymerases may incorporate modified or altered versions of these nucleotides. The order in which the nucleotides are added is dictated by base pairing to a DNA template strand; such base pairing is accomplished through "canonical" hydrogen-bonding (hydrogen-bonding between A and T nucleotides and G and C nucleotides of opposing DNA strands), although non-canonical base pairing, such as G:U base pairing, is known in the art. See e.g., Adams *et al.*, The Biochemistry of the Nucleic Acids 14-32 (11th ed. 1992). The *in-vitro* use of enzymes having DNA polymerase activity has in recent years become more common in a variety of biochemical applications including cDNA synthesis and DNA sequencing reactions (see Sambrook *et al.*, (2nd ed. Cold Spring Harbor Laboratory Press, 1989) hereby incorporated by reference herein), and amplification of nucleic acids by methods such as the polymerase chain reaction (PCR) (Mullis *et al.*, U.S. Pat. Nos. 4,683,195, 4,683,202, and 4,800,159, hereby incorporated by reference herein) and RNA transcription-mediated amplification methods (e.a., Kacian *et al.*, PCT Publication No. WO91/01384).

Methods such as PCR make use of cycles of primer extension through the use of a DNA polymerase activity, followed by thermal denaturation of the resulting double-stranded nucleic acid in order to provide a new template for another round of primer annealing and extension. Because the high temperatures necessary for strand denaturation result in the irreversible inactivations of many DNA polymerases, the discovery and use of DNA polymerases able to remain active at temperatures above about 37°C to 42°C (thermostable DNA polymerase enzymes) provides an advantage in cost and labor efficiency. Thermostable DNA polymerases have been discovered in a number of thermophilic organisms including, but not limited to *Thermus aquaticus*, *Thermus thermophilus*, and species of the *Bacillus*, *Thermococcus*, *Sulfolobus*, *Pyrococcus* genera. DNA polymerases can be purified directly from these thermophilic organisms. However, substantial increases in the yield of DNA polymerase can be obtained by first cloning the gene encoding the enzyme in a multicopy expression vector by recombinant DNA technology methods, inserting the vector into a host cell strain capable of expressing the enzyme, culturing the vector-containing host cells, then extracting the DNA polymerase from a host cell strain which has expressed the enzyme.

The bacterial DNA polymerases that have been characterized to date have certain patterns of similarities and differences which has led some to divide these enzymes into two groups: those whose genes contain introns/inteins (Class B DNA polymerases), and those whose DNA polymerase genes are roughly similar to that of *E. coli* DNA polymerase I and do not contain introns (Class A DNA polymerases).

Several Class A and Class B thermostable DNA polymerases derived from thermophilic organisms have been cloned and expressed. Among the class A enzymes: Lawyer, *et al.*, *J. Biol. Chem.* 264:6427-6437 (1989) and Gelfund *et al.*, U.S. Pat. No. 5,079,352, report the cloning and expression of a full length thermostable DNA polymerase derived from *Thermus aquaticus* (Taq). Lawyer *et al.*, in PCR Methods and Applications, 2:275-287 (1993), and Barnes, PCT Publication No. WO92/06188 (1992), disclose the cloning and expression of truncated versions of the same DNA polymerase, while Sullivan, EPO Publication No. 0482714A1 (1992), reports cloning a mutated version of the Taq DNA polymerase. Asakura *et al.*, *J. Ferment. Bioeng. (Japan)*, 74:265-269 (1993) have reportedly cloned and expressed a DNA polymerase from *Thermus thermophilus*. Gelfund *et al.*, PCT Publication No. WO92/06202 (1992), have disclosed a purified thermostable DNA polymerase from *Thermosipho africanus*. A thermostable DNA polymerase from *Thermus flavus* is reported by Akhmetzjanov and Vakhitov, *Nucleic Acids Res.*, 20:5839 (1992). Uemori *et al.*, *J. Biochem.* 113:401-410 (1993) and EPO Publication No. 0517418A2 (1992) have reported cloning and expressing a DNA polymerase from the thermophilic bacterium *Bacillus caldotenax*. Ishino *et al.*, Japanese Patent Application No. HEI 4[1992]-131400 (publication date Nov. 19, 1993) report cloning a DNA polymerase from *Bacillus stearothermophilus*. Among the Class B enzymes: A recombinant thermostable DNA polymerase from *Thermococcus litoralis* is reported by Comb *et al.*, EPO Publication No. 0 455 430 A3 (1991), Comb *et al.*, EPO Publication No. 0547920A2 (1993), and Perler *et al.*, *Proc. Natl. Acad. Sci. (USA)*, 89:5577-5581 (1992). A cloned thermostable DNA polymerase from *Sulfolobus solfatarius* is disclosed in Pisani *et al.*, *Nucleic Acids Res.* 20:2711-2716 (1992) and in PCT Publication WO93/25691 (1993). The thermostable enzyme of *Pyrococcus furiosus* is disclosed in Uemori *et al.*, *Nucleic Acids Res.*, 21:259-265 (1993), while a recombinant DNA polymerase is derived from *Pyrococcus* sp. as disclosed in Comb *et al.*, EPO Publication No. 0547359A1 (1993).

Many thermostable DNA polymerases possess activities additional to a DNA polymerase activity; these may include a 5'-3' exonuclease activity and/or a 3'-5' exonuclease activity. The activities of 5'-3' and 3'-5' exonucleases are well known to those of ordinary skill in the art. The 3'-5' exonuclease activity improves the accuracy of the newly-synthesized strand by removing incorrect bases that may have been incorporated; DNA polymerases in which such activity is low or absent, reportedly including *Taq* DNA polymerase, (see Lawyer *et al.*, *J. Biol Chem.* 264:6427-6437), have elevated error rates in the incorporation of nucleotide residues into the primer extension strand. In applications such as nucleic acid amplification procedures in which the replication of DNA is often geometric in relation to the number of primer extension cycles, such errors can lead to serious artifactual problems such as sequence heterogeneity of the nucleic acid amplification product (amplicon). Thus, a 3'-5' exonuclease activity is a desired characteristic of a thermostable DNA polymerase used for such purposes.

By contrast, the 5'-3' exonuclease activity often present in DNA polymerase enzymes is often undesired in a particular application since it may digest nucleic acids, including primers, that have an unprotected 5' end. Thus, a thermostable DNA polymerase with an attenuated 5'-3' exonuclease activity, or in which such activity is absent, is also a desired characteristic of an enzyme for biochemical applications. Various DNA polymerase enzymes have been described where a modification has been introduced in a DNA polymerase, which accomplishes this object. For example, the *Klenow* fragment of *E. coli* DNA polymerase I can be produced as a proteolytic fragment of the holoenzyme in which the domain of the protein controlling the 5'-3' exonuclease activity has been removed. The *Klenow* fragment still retains the polymerase activity and the 3'-5' exonuclease activity. Barnes, *supra*, and Gelfund *et al.*, U.S. Pat. No. 5,079,352 have produced 5'-3' exonuclease-deficient recombinant *Taq* DNA polymerases. Ishino *et al.*, EPO Publication No. 0517418A2, have produced a 5'-3' exonuclease-deficient DNA polymerase derived from *Bacillus caldotenax*. On the other hand, polymerases lacking the 5'-3' exonuclease domain often have reduced processivity.

## LIGASE

DNA strand breaks and gaps are generated transiently during replication, repair and recombination. In mammalian cell nuclei, rejoining of such strand breaks depends on several different DNA polymerases and DNA ligase enzymes. The mechanism for joining of DNA strand interruptions by DNA ligase enzymes has been widely described. The reaction is initiated by the formation of a covalent enzyme-adenylate complex. Mammalian and viral DNA ligase enzymes employ ATP as cofactor, whereas bacterial DNA ligase enzymes use NAD to generate the adenylyl group. In the case of ATP-utilising ligases, the ATP is cleaved to AMP and pyrophosphate with the adenylyl residue linked by a phosphoramidate bond to the  $\epsilon$ -amino group of a specific lysine residue at the active site of the protein (Gumport, R. I., *et al.*, *PNAS*, 68:2559-63 (1971)). Reactivated AMP residue of the DNA ligase-adenylate intermediate is transferred to the 5' phosphate terminus of a single strand break in double stranded DNA to generate a covalent DNA-AMP complex with a 5'-5' phosphoanhydride bond. This reaction intermediate has also been isolated for microbial and mammalian DNA ligase enzymes, but is shorter lived than the adenylylated enzyme. In the final step of DNA ligation, unadenylylated DNA ligase enzymes required for the generation of a phosphodiester bond catalyze displacement of the AMP residue through attack by the adjacent 3'-hydroxyl group on the adenylylated site.

The occurrence of three different DNA ligase enzymes, DNA Ligase I, II and III, is established previously by biochemical and immunological characterization of purified enzymes (Tomkinson, A. E. *et al.*, *J. Biol. Chem.*, 266:21728-21735 (1991) and Roberts, E., *et al.*, *J. Biol. Chem.*, 269:3789-3792 (1994)).

## AMPLIFICATION

The methods of our invention involve the templated amplification of desired nucleic acids. "Amplification" refers to the increase in the number of copies of a particular nucleic acid fragment (or a portion of this) resulting either from an enzymatic chain reaction (such as a polymerase chain reaction, a ligase chain reaction, or a self-sustained sequence replication) or from the replication of all or part of the vector into which it has been cloned. Preferably,

the amplification according to our invention is an exponential amplification, as exhibited by for example the polymerase chain reaction.

Many target and signal amplification methods have been described in the literature, for example, general reviews of these methods in Landegren, U., *et al.*, *Science* 242:229-237 (1988) and Lewis, R., *Genetic Engineering News* 10:1, 54-55 (1990). These amplification methods may be used in the methods of our invention, and include polymerase chain reaction (PCR), PCR *in situ*, ligase amplification reaction (LAR), ligase hybridization, Q bacteriophage replicase, transcription-based amplification system (TAS), genomic amplification with transcript sequencing (GAWTS), nucleic acid sequence-based amplification (NASBA) and *in situ* hybridization.

#### Polymerase Chain Reaction (PCR)

PCR is a nucleic acid amplification method described *inter alia* in U.S. Pat. Nos. 4,683,195 and 4,683,202. PCR consists of repeated cycles of DNA polymerase generated primer extension reactions. The target DNA is heat denatured and two oligonucleotides, which bracket the target sequence on opposite strands of the DNA to be amplified, are hybridized. These oligonucleotides become primers for use with DNA polymerase. The DNA is copied by primer extension to make a second copy of both strands. By repeating the cycle of heat denaturation, primer hybridization and extension, the target DNA can be amplified a million fold or more in about two to four hours. PCR is a molecular biology tool, which must be used in conjunction with a detection technique to determine the results of amplification. An advantage of PCR is that it increases sensitivity by amplifying the amount of target DNA by 1 million to 1 billion fold in approximately 4 hours.

The polymerase chain reaction may be used in the selection methods of our invention as follows. For example, PCR may be used to select for variants of *Taq* polymerase having polymerase activity. As described in further detail above, a library of nucleic acids each encoding a replicase or a variant of the replicase, for example, *Taq* polymerase, is generated and subdivided into compartments. Each compartment comprises substantially one member of the library together with the replicase or variant encoded by that member.

The polymerase or variant may be expressed *in vivo* within a transformed bacterium or any other suitable expression host, for example yeast or insect or mammalian cells, and the expression host encapsulated within a compartment. Heat or other suitable means is applied to disrupt the host and to release the polymerase variant and its encoding nucleic acid within the  
5 compartment. In the case of a bacterial host, timed expression of a lytic protein, for example protein E from  $\Phi$ X174, or use of an inducible  $\lambda$  lysogen, may be employed for disrupting the bacterium.

It will be clear that the polymerase or other enzyme need not be a heterologous protein expressed in that host (e.g., a plasmid), but may be expressed from a gene forming part of the  
10 host genome. Thus, the polymerase may be for example an endogenous or native bacterial polymerase. We have shown that in the case of nucleotide diphosphate kinase (ndk), endogenous (uninduced) expression of ndk is sufficient to generate dNTPs for its own replication. Thus, the methods of selection according to our invention may be employed for the direct functional cloning of polymerases and other enzymes from diverse (and uncultured)  
15 microbial populations.

Alternatively, the nucleic acid library may be compartmentalised together with components of an *in vitro* transcription/translation system (as described in further detail in this document), and the polymerase variant expressed *in vitro* within the compartment.

Each compartment also comprises components for a PCR reaction, for example,  
20 nucleotide triphosphates (dNTPs), buffer, magnesium, and oligonucleotide primers. The oligonucleotide primers may have sequences corresponding to sequences flanking the polymerase gene (i.e., within the genomic or vector DNA) or to sequences within the polymerase gene. PCR thermal cycling is then initiated to allow any polymerase variant having polymerase activity to amplify the nucleic acid sequence.

25 Active polymerases will amplify their corresponding nucleic acid sequences, while nucleic acid sequences encoding weakly active or inactive polymerases will be weakly replicated or not be replicated at all. In general, the final copy number of each member of the nucleic acid library will be expected to be proportional to the level of activity of the



polymerase variant encoded by it. Nucleic acids encoding active polymerases will be over-represented, and nucleic acids encoding inactive or weakly active polymerases will be under-represented. The resulting amplified sequences may then be cloned and sequenced, etc., and replication ability of each member assayed.

- 5 As described in further detail elsewhere, the conditions within each compartment may be altered to select for polymerases active under these conditions. For example, heparin may be added to the reaction mix to choose polymerases, which are resistant to heparin. The temperature at which PCR takes place may be elevated to select for heat resistant variants of polymerase. Furthermore, polymerases may be selected which are capable of extending DNA  
10 sequences such as primers with altered 3' ends or altered parts of the primer sequence. The altered 3' ends or other alterations can include unnatural bases (altered sugar or base moieties), modified bases (e.g. blocked 3' ends) or even primers with altered backbone chemistries (e.g. PNA primers).

#### Reverse transcriptase-PCR

- 15 RT-PCR is used to amplify RNA targets. In this process, the reverse transcriptase enzyme is used to convert RNA to complementary DNA (cDNA), which can then be amplified using PCR. This method has proven useful for the detection of RNA viruses.

- The methods of our invention may employ RT-PCR. Thus, the pool of nucleic acids encoding the replicase or its variants may be provided in the form of an RNA library. This  
20 library could be generated *in vivo* in bacteria, mammalian cells, yeast etc., which are compartmentalised, or by *in-vitro* transcription of compartmentalised DNA. The RNA could encode a co-compartmentalised replicase (e.g. reverse transcriptase or polymerase) that has been expressed *in vivo* (and released in emulsion along with the RNA by means disclosed below) or *in vitro*. Other components necessary for amplification (polymerase and/or reverse  
25 transcriptase, dNTPs, primers) are also compartmentalised. Under given selection pressure(s), the cDNA product of the reverse transcription reaction serves as a template for PCR amplification. As with other replication reactions (in particular ndk in the Examples) the RNA may encode a range of enzymes feeding the reaction.

### Self-Sustained Sequence Replication (3SR)

Self-sustained sequence replication (3SR) is a variation of TAS, which involves the isothermal amplification of a nucleic acid template via sequential rounds of reverse transcriptase (RT), polymerase and nuclease activities that are mediated by an enzyme cocktail and appropriate oligonucleotide primers (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874). Enzymatic degradation of the RNA of the RNA/DNA heteroduplex is used instead of heat denaturation. RNase H and all other enzymes are added to the reaction and all steps occur at the same temperature and without further reagent additions. Following this process, amplifications of  $10^6$  to  $10^9$  have been achieved in one hour at 42°C.

10

The methods of our invention may therefore be extended to select polymerases or replicases from mesophilic organisms using 3SR isothermal amplification (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:7797; Compton (1991) *Nature* 350:91-92) instead of PCR thermocycling. As described above, 3SR involves the concerted action of two enzymes: an RNA polymerase as well as a reverse transcriptase cooperate in a coupled reaction of transcription and reverse transcription, leading to the simultaneous amplification of both RNA and DNA. Clearly, in this system self-amplification may be applied to either of the two enzymes involved or to both simultaneously. It may also include the evolution of the RNase H activity either as part of the reverse transcriptase enzyme (e.g. HIV-1 RT) or on its own.

20

The various enzymatic activities that define 3SR and related methods are all targets for selection using the methods of our invention. Variants of either T7 RNA polymerase, reverse transcriptase (RT), or RNaseH can be provided within the aqueous compartments of the emulsions, and selected for under otherwise limiting conditions. These variants can be introduced via *E. coli* "gene pellets" (i.e., bacteria express the polypeptide), or other means as described else where in this document. Initial release in emulsion may be mediated by enzymatic (for example, lambda lysogen) or thermal lysis, or other methods as disclosed here. The latter may necessitate the use of agents that stabilize enzymatic activity at transiently elevated temperatures. For example, it may be necessary to include amounts of proline, glycerol, trehalose or other stabilising agents as known in the art to effect stabilisation of

25

thermosensitive enzymes such as reverse transcriptase. Furthermore, stepwise removal of the agent may be undertaken to select for increased stability of the thermosensitive enzyme.

Alternatively, and as disclosed elsewhere, variants may be produced via coupled transcription translation, with the expressed products feeding into the 3SR cycle.

- 5 It will also be appreciated that it is possible to replace reverse transcriptase with the thermostable Tth DNA polymerase. Tth DNA polymerase is known to have reverse transcriptase activity and the RNA template is effectively reverse-transcribed into template DNA using this enzyme. It is therefore possible to select for useful variants of this enzyme, by for example, introducing bacterially expressed T7 RNA polymerase variants into emulsion  
10 and preincubation at an otherwise non-permissive temperature.

Example 18 below is an example showing one way in which the methods of our invention may be applied to selection of replicases using self-sustained sequence replication (3SR).

#### Ligation Amplification (LAR/LAS)

- 15 Ligation amplification reaction or ligation amplification system uses DNA ligase and four oligonucleotides, two per target strand. This technique is described by Wu, D. Y. and Wallace, R. B. (1989) *Genomics* 4:560. The oligonucleotides hybridize to adjacent sequences on the target DNA and are joined by the ligase. The reaction is heat denatured and the cycle repeated.
- 20 By analogy to the application to polymerases, our method may be applied to ligases in particular from thermophilic organisms. Oligonucleotides complementary to one strand of the ligase gene sequence are synthesized (either as perfect match or comprising targeted or random diversity). The two end oligos overlap into the vector or untranslated regions of the ligase gene. The ligase gene is either cloned for expression in an appropriate host and  
25 compartmentalized together with the oligonucleotides and an appropriate energy source (usually ATP (or NADPH)). If necessary, the ligase expressed as above in bacteria is released

from the cells by thermal lysis. Compartments contain appropriate buffer together with appropriate amounts of an appropriate energy source (ATP or NADH) and oligonucleotides encoding the whole of the ligase gene as well as flanking sequences required for cloning. Ligation of oligonucleotides leads to assembly of a full-length ligase gene (templated by the  
5 ligase gene on the expression plasmid) by an active ligase. In compartments containing an inactive ligase, no assembly will take place. As with polymerases, the copy number of a ligase gene X after self-ligation will preferably be proportional to the catalytic activity under the selection conditions of the ligase X it encodes.

After lysis of the cell, thermocycling leads to annealing of the oligonucleotides to the  
10 ligase gene. However, ligation of the oligos and thus assembly of the full-length ligase gene depends on the presence of an active ligase in the same compartment. Thus only genes encoding active ligases will assemble their own encoding genes from the present oligonucleotides. Assembled genes can then be amplified, diversified and recloned for another round of selection if necessary. The methods of our invention are therefore suitable for the  
15 selection of ligases, which are faster or more efficient at ligation.

As noted elsewhere, the ligase can be produced either *in situ* by expression from a suitable bacterial or other host, or by *in vitro* translation. The ligase may be an oligonucleotide (e.g. ribo or deoxiribozyme) ligase assembling its own sequence from available fragments, or the ligase may be a conventional (polypeptide) ligase. The length of the oligonucleotides will  
20 depend on the particular reaction, but if necessary, they can be very short (e.g. triplets). As noted elsewhere, the method of our invention may be used to select for an agent capable of modulating ligase activity, either directly or indirectly. For example, the gene to be evolved may be another enzyme or enzymes that generates a substrate for the ligase (e.g. NADH) or consumes an inhibitor. In this case the oligonucleotides encode parts of the other enzyme or  
25 enzymes etc.

The ligation reaction between oligonucleotides may incorporate alternative chemistries e.g. amide linkages. As long as the chemical linkages do not interfere with templated copying of the opposite strand by any replicase (e.g. reverse transcriptase), a wide variety of linkage chemistries and ligases that catalyse it may be evolved.

### Q $\beta$ Replicase

In this technique, RNA replicase for the bacteriophage Q $\beta$ , which replicates single-stranded RNA, is used to amplify the target DNA, as described by Lizardi *et al.* (1988) *Bio. Technology* 6:1197. First, the target DNA is hybridized to a primer including a T7 promoter and a Q $\beta$  5' sequence region. Using this primer, reverse transcriptase generates a cDNA  
5 connecting the primer to its 5' end in the process. These two steps are similar to the TAS protocol. The resulting heteroduplex is heat denatured. Next, a second primer containing a Q $\beta$  3' sequence region is used to initiate a second round of cDNA synthesis. This results in a double stranded DNA containing both 5' and 3' ends of the Q $\beta$  bacteriophage as well as an  
10 active T7 RNA polymerase binding site. T7 RNA polymerase then transcribes the double-stranded DNA into new RNA, which mimics the Q $\beta$ . After extensive washing to remove any unhybridized probe, the new RNA is eluted from the target and replicated by Q $\beta$  replicase. The latter reaction creates 10<sup>7</sup> fold amplification in approximately 20 minutes. Significant background may be formed due to minute amounts of probe RNA that is non-specifically  
15 retained during the reaction.

A reaction employing Q $\beta$  replicase as described above may be used to build a continuous selection reaction in an alternative embodiment according to our invention.

For example, the gene for Q $\beta$  replicase (with appropriate 5' and 3' regions) is added to an *in vitro* translation reaction and compartmentalised. In compartments, the replicase is  
20 expressed and immediately starts to replicate its own gene. Only genes encoding an active replicase replicate themselves. Replication proceeds until NTPs are exhausted. However, as NTPs can be made to diffuse through the emulsion (see the description of ndk in the Examples), the replication reaction may be "fed" from the outside and proceed much longer, essentially until there is no room left within the compartments for further replication. It is  
25 possible to propagate the reaction further by serial dilution of the emulsion mix into a fresh oil-phase and re-emulsification after addition of a fresh water-phase containing NTPs. Q $\beta$  replicase is known to be very error-prone, so replication alone will introduce lots of random diversity (which may be desirable). The methods described here allow the evolution of more

specific (e.g. primer dependent) forms of Q $\beta$ -replicase. As with other replication reactions (in particular ndk in the Examples) a range of enzymes feeding the reaction may be evolved.

#### Other Amplification Techniques

Alternative amplification technology may be exploited in the present invention. For example, rolling circle amplification (Lizardi *et al.*, (1998) *Nat Genet* 19:225) is an amplification technology available commercially (RCAT<sup>TM</sup>) which is driven by DNA polymerase and can replicate circular oligonucleotide probes with either linear or geometric kinetics under isothermal conditions.

In the presence of two suitably designed primers, a geometric amplification occurs via DNA strand displacement and hyperbranching to generate  $10^{12}$  or more copies of each circle in 1 hour.

If a single primer is used, RCAT generates in a few minutes a linear chain of thousands of tandemly linked DNA copies of a target covalently linked to that target.

A further technique, strand displacement amplification (SDA; Walker *et al.*, (1992) *PNAS* (USA) 80:392) begins with a specifically defined sequence unique to a specific target. But unlike other techniques which rely on thermal cycling, SDA is an isothermal process that utilizes a series of primers, DNA polymerase and a restriction enzyme to exponentially amplify the unique nucleic acid sequence.

SDA comprises both a target generation phase and an exponential amplification phase.

In target generation, double-stranded DNA is heat denatured creating two single-stranded copies. A series of specially manufactured primers combine with DNA polymerase (amplification primers for copying the base sequence and bumper primers for displacing the newly created strands) to form altered targets capable of exponential amplification.

The exponential amplification process begins with altered targets (single-stranded partial DNA strands with restricted enzyme recognition sites) from the target generation phase.

An amplification primer is bound to each strand at its complimentary DNA sequence.  
5 DNA polymerase then uses the primer to identify a location to extend the primer from its 3' end, using the altered target as a template for adding individual nucleotides. The extended primer thus forms a double-stranded DNA segment containing a complete restriction enzyme recognition site at each end.

A restriction enzyme is then bound to the double stranded DNA segment at its  
10 recognition site. The restriction enzyme dissociates from the recognition site after having cleaved only one strand of the double-sided segment, forming a nick. DNA polymerase recognizes the nick and extends the strand from the site, displacing the previously created strand. The recognition site is thus repeatedly nicked and restored by the restriction enzyme and DNA polymerase with continuous displacement of DNA strands containing the target  
15 segment.

Each displaced strand is then available to anneal with amplification primers as above. The process continues with repeated nicking, extension and displacement of new DNA strands, resulting in exponential amplification of the original DNA target.

#### SELECTION OF CATALYTIC RNA

20 Known methods of *in-vitro* evolution have been used to generate catalytically active RNA molecules (ribozymes) with a diverse range of activities. However, these have involved selection by self-modification, which inherently isolates variants that rely on proximity catalysis and which display reduced activities in trans.

25 Compartmentalisation affords a means to select for truly trans-acting ribozymes capable of multiple turnover, without the need to tether substrate to the ribozyme by covalent linkage or hydrogen-bonding (i.e., base-pairing) interactions.

In its simplest case, a gene encoding a ribozyme can be introduced into emulsion and readily transcribed as demonstrated by the transcription and the 3SR amplification of the RNA encoding Taq polymerase *in situ* as follows: The Taq polymerase gene is first transcribed in emulsion. 100µl of a reaction mix comprising 80mM HEPES-KOH (pH 7.5), 24mM MgCl<sub>2</sub>, 2mM spermidine, 40mM DTT, rNTPs (30mM), 50ng T7-Ta<sub>q</sub> template (see Example 18. Selection Using Self-Sustained Sequence Replication (3SR)), 60 units T7 RNA polymerase (USB), 40 units RNAsin (Promega) is emulsified using the standard protocol. Emulsions are incubated at 37°C for up to 6 hours and analysis of reaction products by gel electrophoresis showed levels of RNA production to be comparable to those of the non-emulsified control.

By creating a 5' overhang (e.g. by ligation of either DNA or RNA adaptors) in the emulsified gene, RNA variants are selected for with the ability of carrying out the template directed addition of successive dNTPs in *trans* (i.e. polymerase activity, see Figure 6). Genes that have been "filled-in" may be rescued by PCR using primers complimentary to the single-stranded region of the gene (i.e., the region, which is single stranded prior to ribozyme fill-in) or by capture of biotin (or otherwise) modified nucleotides that are incorporated followed by PCR. In compartments without catalytic RNA activity, this region remains single stranded, and PCR will fail to amplify the template (alternatively no nucleotides are incorporated and the template is not captured but washed away).

A coupling approach can also be used to further extend the range of enzymatic activities that could be selected for. For example, co-emulsification of a DNA polymerase with the gene described above (5' overhang) can be used to select for ribozymes that convert an otherwise unsuitable NTP substrate into one that can be utilised by the polymerase. As before, the "filled-in" gene can then be rescued by PCR. The above approach can also be used to select for protein polymerase enzyme produced in-situ from a similar template (i.e. with 3' overhang). A diagram showing the selection of RNA having catalytic activity is shown as Figure 6.



## SELECTION OF AGENTS CAPABLE OF MODIFYING REPLICASE ACTIVITY

In another embodiment, our invention is used to select for an agent capable of modifying the activity of a replicase. In this embodiment, a pool of nucleic acids is generated comprising members encoding one or more candidate agents. Members of the nucleic acid library are compartmentalised together with a replicase (which, as explained above, is able only to act on the nucleic acid encoding the agent).

The candidate agents may be functionally or chemically distinct from each other, or they may be variants of an agent known or suspected to be capable of modulating replicase activity. Members of the pool are then segregated into compartments together with the polypeptides or polynucleotides encoded by them, so that preferably each compartment comprises a single member of the pool together with its cognate encoded polypeptide. Each compartment also comprises one or more molecules of the replicase. Thus, the encoded polypeptide agent is able to modulate the activity of the replicase, to prevent or enhance replication of the compartmentalised nucleic acid (i.e., the nucleic acid encoding the agent). In this way, the polypeptide agent is able to act via the replicase to increase or decrease the number of molecules of its encoding nucleic acid. In a highly preferred embodiment of the invention, the agent is capable of enhancing replicase activity, to enable detection or selection of the agent by detecting the encoding nucleic acid.

The modulating agent may act directly or indirectly on the replicase. For example, the modulating agent may be an enzyme comprising an activity, which acts on the replicase molecule, for example, by a post-translational modification of replicase, to activate or inactivate the replicase. The agent may act by taking off or putting on a ligand from the replicase molecule. It is known that many replicases such as polymerases and ligases are regulated by phosphorylation, so that in preferred embodiments the agent according to the invention is a kinase or a phosphorylase. The modulating agent may also directly interact with the replicase and modify its properties (e.g. Thioredoxin & T7-DNA polymerase, members of the replisome e.g. clamp, helicase etc. with DNA polymerase III).

Alternatively, the modulating agent may exert its effects on the replicase in an indirect manner. For example, modulation of replicase activity may take place via a third body, which third body is modified by the modulating agent, for example as described above.

Furthermore, the modulating agent may be an enzyme, which forms part of a pathway, which produces as an end product a substrate for the replicase. In this embodiment, the modulating agent is involved in the synthesis of an intermediate (or the end product) of the pathway. Accordingly, the rate of replication (and hence the amount of nucleic acid encoding the agent) is dependent on the activity of the modulating agent.

For example, the modulating agent may be a kinase that is involved in the biosynthesis of bases, deoxyribonucleosides, deoxyribonucleotides such as dAMP, dCMP, dGMP and dTMP, deoxyribonucleoside diphosphates (such as dADP, dCDP, dCTP and dTDP), deoxyribonucleoside triphosphates such as dATP, dCTP, dGTP or dTTP, or nucleosides, nucleotides such as AMP, CMP, GMP and UMP, nucleoside diphosphates (such as ADP, CDP, CTP and UDP), nucleoside triphosphates such as ATP, CTP, GTP or UTP, etc. The modulating agent may be involved in the synthesis of other intermediates in the biosynthesis of nucleotides (as described and well known from biochemical textbooks such as Stryer or Lehninger), such as IMP, 5-phospho- $\alpha$ -D-ribose-1-pyrophosphoric acid, 5-phospho- $\beta$ -D-ribose-5-phosphate, 5-phosphoribosyl-glycinamide, 5-phosphoribosyl-N-formylglycinamide, etc. Thus, the agent may comprise an enzyme such as ribosephosphate pyrophosphokinase, phosphoribosylglycinamide synthetase, etc. Other examples of such agents will be apparent to those skilled in the art. The methods of our invention allow the selection of such agents with improved catalytic activity.

In yet another embodiment, the modulator functions to "unblock" a constituent of the replication cocktail (primers, dNTP, replicase etc). An example of a blocked constituent would be a primer or dNTP with a chemical moiety attached that inhibits the replicase used in the CSR cycle. Alternatively, the pair of primers used could be covalently tethered by a linking agent, with cleavage of the agent by the modulator allowing both primers to amplify its gene in the presence of supplemented replicase. An example of a linking agent would be a peptide nucleic acid (PNA). Additionally, by designing a large oligonucleotide that encodes a

pair of primer sequences interspersed by target nucleotide sequence, novel site-specific restriction enzymes could be evolved. As before, the rate of replication (and hence the amount of nucleic acid encoding the agent) is dependent on the activity of the modulating agent. Alternatively the modulator can modify the 5' end a primer such that amplification products  
5 incorporating the primer can be captured by a suitable agent (e.g. antibody) and thus enriched and reamplified.

In a further embodiment, the scope of CSR may be further broadened to select for agents that are not necessarily thermostable. Delivery vehicles (e.g. *E.coli*) containing expression constructs that encode a secretable form of a modulator/replicase of interest are  
10 compartmentalised. Inclusion of an inducing agent in the aqueous phase and incubation at permissive temperature (e.g. 37°C) allows for expression and secretion of the modulator/replicase into the compartment. Sufficient time is then allowed for the modulator to act in any of the aforementioned ways to facilitate subsequent amplification of the gene encoding it (e.g. consume an inhibitor of replication). The ensuing temperature change during  
15 the amplification process serves to rid the compartment of host cell enzymatic activities (that have up to this point been segregated from the aqueous phase) and release the encoding gene for amplification.

Thus, according to an embodiment of our invention, we provide a method of selecting a polypeptide involved in a pathway which has as an end product a substrate which is  
20 involved in a replication reaction ("a pathway polypeptide"), the method comprising the steps of: (a) providing a replicase; (b) providing a pool of nucleic acids comprising members each encoding a pathway polypeptide or a variant of the pathway polypeptide; (c) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool, the pathway polypeptide or variant encoded by the nucleic acid  
25 member, the replicase, and other components of the pathway; and (d) detecting amplification of the nucleic acid member by the replicase.

The Examples (in particular Example 19 and following Examples) show the use of our invention in the selection of nucleoside diphosphate kinase (NDP Kinase), which catalyses the

transfer of a phosphate group from ATP to a deoxynucleoside diphosphate to produce a deoxynucleoside triphosphate.

In yet another embodiment, the modulating agent is such that it consumes an inhibitor of replicase activity. For example, it is known that heparin is an inhibitor of replicase (polymerase) activity. Our method allows the selection of a heparinase with enhanced activity, by compartmentalisation of a library of nucleic acids encoding heparinase or variants of this enzyme, in the presence of heparin and polymerase. Heparinase variants with enhanced activity are able to break down heparin to a greater extent or more rapidly, thus removing the inhibition of replicase activity within the compartment and allowing the replication of the nucleic acid within the compartment (i.e., the nucleic acid encoding that heparinase variant).

#### SELECTION OF INTERACTING POLYPEPTIDES

The most important systems for the selection of protein-protein interactions are *in vivo* methods, with the most important and best developed being the yeast two-hybrid system (Fields & Song, *Nature* (1989) 340, 245-246). In this system and related approaches two hybrid proteins are generated: a bait-hybrid comprising protein X fused to a DNA-binding domain and a prey-hybrid comprising protein Y fused to a transcription activation domain with cognate interaction of X and Y reconstituting the transcriptional activator. Two other *in vivo* systems have been put forward in which the polypeptide chain of an enzyme is expressed in two parts fused to two proteins X and Y and in which cognate X-Y interaction reconstitutes function of the enzyme (Karimova (1998) *Proc Natl Acad Sci U S A*, 95, 5752-6; Pelletier (1999) *Nat Biotechnol*, 17, 683-690) conferring a selectable phenotype on the cell.

It has recently been shown that Taq polymerase can be split in a similar way (Vainshtein *et al* (1996) *Protein Science* 5, 1785). According to our invention, therefore, we provide a method of selecting a pair of polypeptides capable of stable interaction by splitting Taq polymerase or any enzyme or factor auxiliary to the polymerase reaction.

The method comprises several steps. The first step consists of providing a first nucleic acid and a second nucleic acid. The first nucleic acid encodes a first fusion protein comprising

a first subdomain of a replicase (or other see above) enzyme fused to a first polypeptide, while the second nucleic acid encodes a second fusion protein comprising a second subdomain of a replicase (or other see above) enzyme fused to a second polypeptide. The two fusion proteins are such that stable interaction of the first and second replicase (or other see above) subdomains generates replicase activity (either directly or indirectly). At least one of the first and second nucleic acids (preferably both) is provided in the form of a pool of nucleic acids encoding variants of the respective first and/or second polypeptide(s).

The pool or pools of nucleic acids are then subdivided into compartments, such that each compartment comprises a first nucleic acid and a second nucleic acid together with respective fusion proteins encoded by the first and second nucleic acids. The first polypeptide is then allowed to bind to the second polypeptide, such that binding of the first and second polypeptides leads to stable interaction of the replicase subdomains to generate replicase activity. Finally, amplification of at least one of the first and second nucleic acids by the replicase is detected

Our invention therefore encompasses an *in vitro* selection system whereby reconstitution of replicase function through the cognate association of two polypeptide ligands drives amplification and linkage of the genes of the two ligands. Such an *in vitro* two-hybrid system is particularly suited for the investigation of protein-protein interactions at high temperatures, e.g. for the investigation of the proteomes of thermophilic organisms or the engineering of highly stable interactions.

The system can also be applied to the screening and isolation of molecular compounds that promote cognate interactions. For example, compounds can be chemically linked to either primers or dNTPs and thus would only be incorporated into amplicons if promoting association. In order to prevent cross-over, such compounds would have to be released only after compartmentalisation has taken place, e.g. by coupling to microbeads or by inclusion into dissolvable microspheres.

## SINGLE STEP AND MULTIPLE STEP SELECTIONS

The selection of suitable encapsulation conditions is desirable. Depending on the complexity and size of the library to be screened, it may be beneficial to set up the encapsulation procedure such that 1 or less than 1 nucleic acids is encapsulated per microcapsule or compartment. This will provide the greatest power of resolution. Where the library is larger and/or more complex, however, this may be impracticable; it may be preferable to encapsulate or compartmentalise several nucleic acids together and rely on repeated application of the method of the invention to achieve sorting of the desired activity. A combination of encapsulation procedures may be used to obtain the desired enrichment.

Theoretical studies indicate that the larger the number of nucleic acids variants created the more likely it is that a molecule will be created with the properties desired (see Perelson and Oster, 1979 *J Theor Biol*, 81, 64570 for a description of how this applies to repertoires of antibodies). Recently it has also been confirmed practically that larger phage-antibody repertoires do indeed give rise to more antibodies with better binding affinities than smaller repertoires (Griffiths *et al.*, (1994) *Embo J*, 13, 3245-60). To ensure that rare variants are generated and thus are capable of being selected, a large library size is desirable. Thus, the use of optimally small microcapsules is beneficial.

In addition to the nucleic acids described above, the microcapsules or compartments according to the invention may comprise further components required for the replication reaction to take place. Other components of the system may for example comprise those necessary for transcription and/or translation of the nucleic acid. These are selected for the requirements of a specific system from the following; a suitable buffer, an *in vitro* transcription/replication system and/or an *in vitro* translation system containing all the necessary ingredients, enzymes and cofactors, RNA polymerase, nucleotides, nucleic acids (natural or synthetic), transfer RNAs, ribosomes and amino acids, and the substrates of the reaction of interest in order to allow selection of the modified gene product.

### Buffer

A suitable buffer will be one in which all of the desired components of the biological system are active and will therefore depend upon the requirements of each specific reaction system. Buffers suitable for biological and/or chemical reactions are known in the art and  
5 recipes provided in various laboratory texts (Sambrook *et al.*, (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York).

### In vitro Translation

The replicase may be provided by expression from a suitable host as described elsewhere, or it may be produced by *in vitro* transcription/translation in a suitable system as  
10 known in the art.

The *in vitro* translation system will usually comprise a cell extract, typically from bacteria (Zubay, 1973, *Annu Rev Genet*, 7, 267-87; Zubay, 1980, *Methods Enzymol*, 65, 856-77; Lesley *et al.*, 1991 *J Biol Chem* 266(4), 2632-8; Lesley, 1995 *Methods Mol Biol*, 37, 265-78.), rabbit reticulocytes (Pelham and Jackson, 1976, *Eur J Biochem*, 67, 247-56), or wheat  
15 germ (Anderson *et al.*, 1983, *Methods Enzymol*, 101, 635-44). Many suitable systems are commercially available (for example from Promega) including some which will allow coupled transcription/translation (all the bacterial systems and the reticulocyte and wheat germ TNT<sup>TM</sup> extract systems from Promega). The mixture of amino acids used may include synthetic amino acids if desired, to increase the possible number or variety of proteins produced in the library.  
20 This can be accomplished by charging tRNAs with artificial amino acids and using these tRNAs for the *in vitro* translation of the proteins to be selected (Ellman *et al.*, 1991, *Methods Enzymol*, 202, 301-36; Benner, 1994, *Trends Biotechnol*, 12, 158-63; Mendel *et al.*, 1995, *Annu Rev Biophys Biomol Struct*, 24, 435-62). Particularly desirable may be the use of *in vitro* translation systems reconstituted from purified components like the PURE system (Shimizu *et al* (2001) *Nat. Biotech.*, 19, 751).  
25

After each round of selection the enrichment of the pool of nucleic acids for those encoding the molecules of interest can be assayed by non-compartmentalised *in vitro*

transcription/replication or coupled transcription-translation reactions. The selected pool is cloned into a suitable plasmid vector and RNA or recombinant protein is produced from the individual clones for further purification and assay.

5 The invention moreover relates to a method for producing a gene product, once a nucleic acid encoding the gene product has been selected by the method of the invention. Clearly, the nucleic acid itself may be directly expressed by conventional means to produce the gene product. However, alternative techniques may be employed, as will be apparent to those skilled in the art. For example, the genetic information incorporated in the gene product may be incorporated into a suitable expression vector, and expressed therefrom.

## 10 COMPARTMENTS

As used here, the term "compartment" is synonymous with "microcapsule" and the terms are used interchangeably. The function of the compartment is to enable co-localisation of the nucleic acid and the corresponding polypeptide encoded by the nucleic acid. This is preferably achieved by the ability of the compartment to substantially restrict diffusion of  
15 template and product strands to other compartments. Any replicase activity of the polypeptide is therefore restricted to being exercised on a nucleic acid within the confines of a compartment, and not other nucleic acids in other compartments. Another function of compartments is to restrict diffusion of molecules generated in a chemical or enzymatic reaction that feed or unblock a replication reaction.

20 The compartments of the present invention therefore require appropriate physical properties to allow the working of the invention.

First, to ensure that the nucleic acids and polypeptides do not diffuse between compartments, the contents of each compartment must be isolated from the contents of the surrounding compartments, so that there is no or little exchange of the nucleic acids and  
25 polypeptides between the compartments over a significant timescale.



Second, the method of the present invention requires that there are only a limited number of nucleic acids per compartment, or that all members within a single compartment are clonal (i.e. identical). This ensures that the polypeptide encoded by and corresponding to an individual nucleic acid will be isolated from other different nucleic acids. Thus, coupling  
5 between nucleic acid and its corresponding polypeptide will be highly specific. The enrichment factor is greatest with on average one or fewer nucleic acid clonal species per compartment, the linkage between nucleic acid and the activity of the encoded polypeptide being as tight as is possible, since the polypeptide encoded by an individual nucleic acid will be isolated from the products of all other nucleic acids. However, even if the theoretically  
10 optimal situation of, on average, a single nucleic acid or less per compartment is not used, a ratio of 5, 10, 50, 100 or 1000 or more nucleic acids per compartment may prove beneficial in selecting from a large library. Subsequent rounds of selection, including renewed compartmentalisation with differing nucleic acid distribution, will permit more stringent selection of the nucleic acids. Preferably, on average there is a single nucleic acid clonal  
15 species, or fewer, per compartment.

Moreover, each compartment contains a nucleic acid; this means that whilst some compartments may remain empty, the conditions are adjusted such that, statistically, each compartment will contain at least one, and preferably only one, nucleic acid.

Third, the formation and the composition of the compartments must not abolish the  
20 function of the machinery for the expression of the nucleic acids and the activity of the polypeptides.

Consequently, any compartmentalisation system used must fulfil these three requirements. The appropriate system(s) may vary depending on the precise nature of the requirements in each application of the invention, as will be apparent to the skilled person.

25 Various technologies are available for compartmentalisation, for example, gas aphrons (Juaregi and Varley, 1998, *Biotechnol Bioeng* 59, 471 and prefabricated nanowells (Huang and Schreiber, 1997, *Proc. Natl. Acad. Sci USA*, 94, 25). For different applications, different compartment sizes and surface chemistries, as discussed in further detail below, may be

desirable. For example, it may be sufficient to utilise diffusion limiting porous materials like gels or alginate (Draget et al., 1997, *Int J Macromol* 21, 47) or zeolithe-type materials. Furthermore, where in-situ PCR or in-cell PCR is carried out, cells may be treated with a cross-linking fixative to form porous compartments allowing diffusion of dNTPs, enzymes and primers.

A wide variety of compartmentalisation or microencapsulation procedures are available (Benita, S., Ed. (1996). *Microencapsulation: methods and industrial applications. Drugs and pharmaceutical sciences.* Edited by Swarbrick, J. New York: Marcel Dekker) and may be used to create the compartments used in accordance with the present invention.

Indeed, more than 200 microencapsulation or compartmentalisation methods have been identified in the literature (Finch, C. A. (1993) *Encapsulation and controlled release. Spec. Publ.-R. Soc. Chem.* 138, 35)

These include membrane enveloped aqueous vesicles such as lipid vesicles (liposomes) (New, R. R. C., Ed. (1990). *Liposomes: a practical approach. The practical approach series.* Edited by Rickwood, D. & Hames, B. D. Oxford: Oxford University Press) and non-ionic surfactant vesicles (van Hal, D. A., Bouwstra, J. A. & Junginger, H. E. (1996). *Nonionic surfactant vesicles containing estradiol for topical application. In Microencapsulation: methods and industrial applications* (Benita, S., ed.), pp. 329-347. Marcel Dekker, New York.). These are closed-membranous capsules of single or multiple bilayers of non-covalently assembled molecules, with each bilayer separated from its neighbour by an aqueous compartment. In the case of liposomes the membrane is composed of lipid molecules; these are usually phospholipids but sterols such as cholesterol may also be incorporated into the membranes (New, R. R. C., Ed. (1990). *Liposomes: a practical approach. The practical approach series.* Edited by Rickwood, D. & Hames, B. D. Oxford: Oxford University Press). A variety of enzyme-catalysed biochemical reactions, including RNA and DNA polymerisation, can be performed within liposomes (Chakrabarti *J Mol Evol.* (1994), 39, 555-9; Oberholzer *Biochem Biophys Res Commun.* (1995), 207, 250-7; Oberholzer *Chem Biol.* (1995) 2, 677-82.; Walde, *Biotechnol Bioeng* (1998), 57, 216-219; Wick & Luisi, *Chem Biol.* (1996), 3, 277-85).

With a membrane-enveloped vesicle system much of the aqueous phase is outside the vesicles and is therefore non-compartmentalised. This continuous, aqueous phase should be removed or the biological systems in it inhibited or destroyed (for example, by digestion of nucleic acids with DNase or RNase) in order that the reactions are limited to the compartmentalised microcapsules (Luisi *et al.*, *Methods Enzymol.* 1987, 136, 188-216).

Enzyme-catalysed biochemical reactions have also been demonstrated in microcapsule compartments generated by a variety of other methods. Many enzymes are active in reverse micellar solutions (Bru & Walde, *Eur J Biochem.* 1991, 199, 95-103.; Bru & Walde, *Biochem Mol Biol Int.* 1993, 31, 685-92; Creagh *et al.*, *Enzyme Microb Technol.* 1993, 15, 383-92; Haber *et al.*, 1993 UNABLE TO FIND; Kumar *et al.*, *Biophys J* 1989, 55, 789-792; Luisi, P. L. & B., S.-H. (1987). Activity and conformation of enzymes in reverse micellar solutions. *Methods Enzymol* 136(188), 188-216; Mao & Walde, *Biochem Biophys Res Commun* 1991, 178, 1105-1112; Mao, Q. & Walde, P. (1991). Substrate effects on the enzymatic activity of alpha-chymotrypsin in reverse micelles. *Biochem Biophys Res Commun* 178(3), 1105-12; Mao *Eur J Biochem.* 1992, 208, 165-70; Perez, G. M., Sanchez, F. A. & Garcia, C. F. (1992). Application of active-phase plot to the kinetic analysis of lipoxygenase in reverse micelles. *Biochem J.*; Walde, P., Goto, A., Monnard, P.-A., Wessicken, M. & Luisi, P. L. (1994) Oparin's reactions revisited: enzymatic synthesis of poly(adenylic acid) in micelles and self-reproducing vesicles. *J. Am. Chem. Soc.* 116, 7541-7547; Walde, P., Han, D. & Luisi, P. L. (1993). Spectroscopic and kinetic studies of lipases solubilized in reverse micelles. *Biochemistry* 32, 4029-34; Walde *Eur J Biochem.* 1988 173, 401-9) such as the AOT-isooctane-water system (Menger, F. M. & Yamada, K. (1979). *J. Am. Chem. Soc.* 101, 6731-6734).

25

Compartments can also be generated by interfacial polymerisation and interfacial complexation (Whateley, T. L. (1996) Microcapsules: preparation by interfacial polymerisation and interfacial complexation and their applications. In *Microencapsulation: methods and industrial applications* (Benita, S., ed.), pp. 349-375. Marcel Dekker, New York). Microcapsule compartments of this sort can have rigid, nonpermeable membranes, or semipermeable membranes. Semipermeable microcapsules bordered by cellulose nitrate membranes, polyamide membranes and lipid-polyamide membranes can all support

30

biochemical reactions, including multienzyme systems (Chang, *Methods Enzymol.* 1987, 136, 67-82; Chang, *Artif Organs.* 1992, 16, 71-4; Lim *Appl Biochem Biotechnol.* 1984, 10:81-5). Alginate/polylysine compartments (Lim & Sun, *Science* (1980) 210, 908-10), which can be formed under very mild conditions, have also proven to be very biocompatible, providing, for example, an effective method of encapsulating living cells and tissues (Chang, *Artif Organs.* (1992) 16, 71-4; Sun *ASAIO J.* (1992), 38, 125-7).

Non-membranous compartmentalisation systems based on phase partitioning of an aqueous environment in a colloidal system, such as an emulsion, may also be used.

Preferably, the compartments of the present invention are formed from emulsions; heterogeneous systems of two immiscible liquid phases with one of the phases dispersed in the other as droplets of microscopic or colloidal size (Becher, P. (1957) *Emulsions: theory and practice.* Reinhold, New York; Sherman, P. (1968) *Emulsion science.* Academic Press, London; Lissant, K.J., ed Emulsions and emulsion technology. Surfactant Science New York: Marcel Dekker, 1974; Lissant, K.J., ed. Emulsions and emulsion technology. Surfactant Science New York: Marcel Dekker, 1984).

Emulsions may be produced from any suitable combination of immiscible liquids. Preferably the emulsion of the present invention has water (containing the biochemical components) as the phase present in the form of finely divided droplets (the disperse, internal or discontinuous phase) and a hydrophobic, immiscible liquid (an 'oil') as the matrix in which these droplets are suspended (the nondisperse, continuous or external phase). Such emulsions are termed 'water-in-oil' (W/O). This has the advantage that the entire aqueous phase containing the biochemical components is compartmentalised in discrete droplets (the internal phase). The external phase, being a hydrophobic oil, generally contains none of the biochemical components and hence is inert.

The emulsion may be stabilised by addition of one or more surface-active agents (surfactants). These surfactants are termed emulsifying agents and act at the water/oil interface to prevent (or at least delay) separation of the phases. Many oils and many emulsifiers can be used for the generation of water-in-oil emulsions; a recent compilation listed over 16,000

surfactants, many of which are used as emulsifying agents (Ash, M. and Ash, I. (1993) Handbook of industrial surfactants. Gower, Aldershot). Suitable oils include light white mineral oil and non-ionic surfactants (Schick, 1966 not found) such as sorbitan monooleate (Span<sup>TM</sup>80; ICI) and polyoxyethylenesorbitan monooleate (Tween<sup>TM</sup> 80; ICI) or t-  
5 Octylphenoxypolyethoxyethanol (Triton X-100).

The use of anionic surfactants may also be beneficial. Suitable surfactants include sodium cholate and sodium taurocholate. Particularly preferred is sodium deoxycholate, preferably at a concentration of 0.5% w/v, or below. Inclusion of such surfactants can in some cases increase the expression of the nucleic acids and/or the activity of the polypeptides.  
10 Addition of some anionic surfactants to a non-emulsified reaction mixture completely abolishes translation. During emulsification, however, the surfactant is transferred from the aqueous phase into the interface and activity is restored. Addition of an anionic surfactant to the mixtures to be emulsified ensures that reactions proceed only after compartmentalisation.

Creation of an emulsion generally requires the application of mechanical energy to  
15 force the phases together. There are a variety of ways of doing this which utilise a variety of mechanical devices, including stirrers (such as magnetic stir-bars, propeller and turbine stirrers, paddle devices and whisks), homogenisers (including rotor-stator homogenisers, high-pressure valve homogenisers and jet homogenisers), colloid mills, ultrasound and 'membrane emulsification' devices (Becher, P. (1957) Emulsions: theory and practice.  
20 Reinhold, New York; Dickinson, E. (1994) In Wedlock, D.J. (ed.), Emulsions and droplet size control. Butterworth-Heine-mann, Oxford, Vol. pp. 191-257).

Aqueous compartments formed in water-in-oil emulsions are generally stable with little if any exchange of polypeptides or nucleic acids between compartments. Additionally, it is known that several biochemical reactions proceed in emulsion compartments. Moreover,  
25 complicated biochemical processes, notably gene transcription and translation are also active in emulsion microcapsules. The technology exists to create emulsions with volumes all the way up to industrial scales of thousands of litres (Becher, P. (1957) Emulsions: theory and practice. Reinhold, New York; Sherman, P. (1968) Emulsion science. Academic Press, London; Lissant, K.J., ed Emulsions and emulsion technology. Surfactant Science New

York: Marcel Dekker, 1974; Lissant, K.J., ed. Emulsions and emulsion technology. Surfactant Science New York: Marcel Dekker, 1984).

5 The preferred compartment size will vary depending upon the precise requirements of any individual selection process that is to be performed according to the present invention. In all cases, there will be an optimal balance between gene library size, the required enrichment and the required concentration of components in the individual compartments to achieve efficient expression and reactivity of the polypeptides.

10 The processes of expression may occur either *in situ* within each individual microcapsule or exogenously within cells (e.g. bacteria) or other suitable forms of subcompartmentalization. Both *in vitro* transcription and coupled transcription-translation become less efficient at sub-nanomolar DNA concentrations. Because of the requirement for only a limited number of DNA molecules to be present in each compartment, this therefore sets a practical upper limit on the possible compartment size where *in vitro* transcription is used. Preferably, for expression *in situ* using *in vitro* transcription and/or translation the mean  
15 volume of the compartments is less than  $5.2 \times 10^{-16} \text{ m}^3$ , (corresponding to a spherical compartment of diameter less than  $1 \mu\text{m}$ ).

An alternative is the separation of expression and compartmentalisation, e.g. using a cellular host. For inclusion of cells (in particular eucaryotic cells) mean compartment diameters of larger than  $10 \mu\text{m}$  may be preferred.

20 As shown in the Examples, to colocalize the polymerase gene and encoded protein within the same emulsion compartment, we used bacteria (*E.coli*) overexpressing *Taq* polymerase as "delivery vehicles". *E.coli* cells (diameter  $1\text{-}5 \mu\text{m}$ ) fit readily into our emulsion compartments while leaving room for sufficient amounts of PCR reagents like nucleotide triphosphates and primers (as shown in Fig. 2). The denaturation step of the first PCR cycle  
25 ruptures the bacterial cell and releases the expressed polymerase and its encoding gene into the compartment allowing self-replication to proceed while simultaneously destroying background bacterial enzymatic activities. Furthermore, by analogy to hot-start strategies, this

cellular "subcompartmentalization" prevents release of polymerase activity at ambient temperatures and the resulting non-specific amplification products.

The effective DNA or RNA concentration in the compartments may be artificially  
5 increased by various methods that will be well-known to those versed in the art. These include, for example, the addition of volume excluding chemicals such as polyethylene glycols (PEG) and a variety of gene amplification techniques, including transcription using RNA polymerases including those from bacteria such as *E. coli* (Roberts, 1969 *Nature* 224, 1168-74; Blattner and Dahlberg, 1972 *Nat New Biol.* 237, 227-32; Roberts *et al.*, 1975 *J Biol*  
10 *Chem.* 250, 5530-41; Rosenberg *et al.*, 1975 *J Biol Chem* 250, 4755-4764), eukaryotes e. g. (Weil *et al.*, 1979 *J Biol Chem.* 254, 6163-6173; Manley *et al.*, 1983 *Methods Enzymol.* 101, 568-82) and bacteriophage such as T7, T3 and SP6 (Melton *et al.*, 1984 *Nucleic Acids Res.* 12, 7035-56.); the polymerase chain reaction (PCR) (Saiki *et al.*, 1988 *Science* 239, 487-91); Q $\beta$  replicase amplification (Miele *et al.*, 1983 *J Mol Biol.* 171, 281-95; Cahill *et al.*, 1991 *Clin*  
15 *Chem.* 37, 1482-5; Chetverin and Spirin, 1995 *Prog Nucleic Acid Res Mol Biol* 51, 225-70; Katanaev *et al.*, 1995 *FEBS Lett.*, 359, 89-92); the ligase chain reaction (LCR) (Landegren *et al.*, 1988 *Science*, 241; 1077-80; Barany, 1991 *PCR Methods Appl.*, 1, 5-16); and self-sustained sequence replication system (Fahy *et al.*, 1991 *PCR Methods Appl.* 1, 25-33) and strand displacement amplification (Walker *et al.*, 1992 *Nucleic Acids Res.* 20, 1691-6).  
20 Gene amplification techniques requiring thermal cycling such as PCR and LCR may also be used if the emulsions and the *in vitro* transcription or coupled transcription-translation systems are thermostable (for example, the coupled transcription-translation systems could be made from a thermostable organism such as *Thermus aquaticus*).

Increasing the effective local nucleic acid concentration enables larger compartments  
25 to be used effectively.

The compartment size must be sufficiently large to accommodate all of the required components of the biochemical reactions that are needed to occur within the compartment. For example, *in vitro*, both transcription reactions and coupled transcription-translation reactions require a total nucleoside triphosphate concentration of about 2mM.

For example, in order to transcribe a gene to a single short RNA molecule of 500 bases in length, this would require a minimum of 500 molecules of nucleoside triphosphate per compartment ( $8.33 \times 10^{-22}$  moles). In order to constitute a 2mM solution, this number of molecules must be contained within a compartment of volume  $4.17 \times 10^{-19}$  litres ( $4.17 \times 10^{-22}$  m<sup>3</sup> which if spherical would have a diameter of 93nm. Hence, the preferred lower limit for microcapsules is a diameter of approximately 0.1µm (100nm).

When using expression hosts as delivery vehicles, there are much less strict requirements on the compartment size. Basically, the compartment has to be of sufficient size to contain the expression host as well as sufficient amounts of reagents to carry out the required reactions. Thus, in such cases larger compartment sizes >10µm are preferred. By an appropriate choice of vector used for expression in the host, the template concentration within compartments can be controlled via the vector origin and resulting copy number (e.g. *E.coli*: colE (pUC) >100, p15: 30-50, pSC101:1-4). Likewise the concentration of the gene product can be controlled by the amount by choice of expression promoter and expression protocol (e.g. full induction of expression versus promoter leakage). Preferably, gene product concentration is as high as possible.

Furthermore, the use of feeder compartments allows feeding of substrates from the outside (see Ghadessy et al .(2001), PNAS, 98, 4552; 01). Feeding emulsion reactions from the outside may allow compartment dimensions <0.1µm for ribozyme selections, as reagents do not need to be contained in their entirety within the compartment.

The size of emulsion microcapsules or compartments may be varied simply by tailoring the emulsion conditions used to form the emulsion according to requirements of the selection system. The larger the compartment size, the larger is the volume that will be required to encapsulate a given nucleic acid library, since the ultimately limiting factor will be the size of the compartment and thus the number of microcapsule compartments possible per unit volume.

The size of the compartments is selected not only having regard to the requirements of the replication system, but also those of the selection system employed for the nucleic acid.



Thus, the components of the selection system, such as a chemical modification system, may require reaction volumes and/or reagent concentrations, which are not optimal for replication. As set forth herein, such requirements may be accommodated by a secondary re-encapsulation step; moreover, they may be accommodated by selecting the compartment size in order to  
5 maximise replication and selection as a whole. Empirical determination of optimal compartment volume and reagent concentration, for example, as set forth herein, is preferred.

In a highly preferred embodiment of the present invention, the emulsion is a water-in-oil emulsion. The water-in-oil emulsion is made by adding an aqueous phase dropwise to an oil phase in the presence of a surfactant comprising 4.5% (v/v) Span 80, about 0.4% (v/v)  
10 Tween 80 and about 0.05-0.1% (v/v) Triton X100 in mineral oil preferably at a ratio of oil:water phase of 2:1 or 3:1. It appears that the ratio of the three surfactants is important for the advantageous properties of the emulsion, and accordingly, our invention also encompasses a water-in-oil emulsion having increased amounts of surfactant but with substantially the same ratio of Span 80, Tween 80 and Triton X100. In a preferred embodiment, the surfactant  
15 comprises 4.5% (v/v) Span 80, 0.4% (v/v) Tween 80 and 0.05% (v/v) Triton X100.

The water-in-oil emulsion is preferably formed under constant stirring in 2ml round bottom biofreeze vials with continued stirring at 1000rpm for a further 4 or 5 minutes after complete addition of the aqueous phase. The rate of addition may be up to 12 drops/min (ca. 10µl each). The aqueous phase may include just water, or it may comprise a buffered solution  
20 having additional components such as nucleic acids, nucleotide triphosphates, etc. In a preferred embodiment, the aqueous phase comprises a PCR reaction mix as disclosed elsewhere in this document, as well as nucleic acid, and polymerase. The water-in-oil emulsion may be formed from 200µl of aqueous phase (for example PCR reaction mix) and 400µl oil phase as described above.

25 The water-in-oil emulsion according to the invention has advantageous properties of increased thermal stability. Thus, no changes in compartment size or evidence of coalescence is observed after 20 cycles of PCR as judged by laser diffraction and light microscopy. This is shown in Figure 2. In addition, polymerase chain reaction proceeded efficiently within the compartments of this water-in-oil composition, to approach the rates observed in solution

PCR. Average aqueous compartment dimensions in the water-in-oil emulsion according to our invention are on average 15 $\mu$ m in size. Once formed, the compartments of the emulsion according to our invention do not permit the exchange of macromolecules like DNA and proteins to any significant degree (as shown in Figure 3A). This is presumably because the  
5 large molecular weight and charged nature of the macromolecules precludes diffusion across the hydrophobic surfactant shell, even at elevated temperatures.

#### NUCLEIC ACIDS

A nucleic acid in accordance with the present invention is as described above. Preferably, the nucleic acid is a molecule or construct selected from the group consisting of a  
10 DNA molecule, an RNA molecule, a partially or wholly artificial nucleic acid molecule consisting of exclusively synthetic or a mixture of naturally-occurring and synthetic bases, any one of the foregoing linked to a polypeptide, and any one of the foregoing linked to any other molecular group or construct. Advantageously, the other molecular group or construct may be selected from the group consisting of nucleic acids, polymeric substances, particularly beads,  
15 for example polystyrene beads, magnetic substances such as magnetic beads, labels, such as fluorophores or isotopic labels, chemical reagents, binding agents such as macrocycles and the like.

The nucleic acid may comprise suitable regulatory sequences, such as those required for efficient expression of the gene product, for example promoters, enhancers, translational  
20 initiation sequences, polyadenylation sequences, splice sites and the like.

The terms "isolating", "sorting" and "selecting", as well as variations thereof, are used herein. Isolation, according to the present invention, refers to the process of separating an entity from a heterogeneous population, for example a mixture, such that it is free of at least one substance with which it is associated before the isolation process. In a preferred  
25 embodiment, isolation refers to purification of an entity essentially to homogeneity. Sorting of an entity refers to the process of preferentially isolating desired entities over undesired entities. In as far as this relates to isolation of the desired entities, the terms "isolating" and "sorting" are equivalent. The method of the present invention permits the sorting of desired

nucleic acids from pools (libraries or repertoires) of nucleic acids which contain the desired nucleic acid. Selecting is used to refer to the process (including the sorting process) of isolating an entity according to a particular property thereof.

“Oligonucleotide” refers to a molecule comprised of two or more  
5 deoxyribonucleotides or ribonucleotides, preferably more than three. The exact size of the oligonucleotide will depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be derived synthetically or by cloning.

The nucleic acids selected according to our invention may be further manipulated. For example, nucleic acid encoding selected replicase or interacting polypeptides are incorporated  
10 into a vector, and introduced into suitable host cells to produce transformed cell lines that express the gene product. The resulting cell lines can then be propagated for reproducible qualitative and/or quantitative analysis of the effect(s) of potential drugs affecting gene product function. Thus gene product expressing cells may be employed for the identification of compounds, particularly small molecular weight compounds, which modulate the function  
15 of gene product. Thus host cells expressing gene product are useful for drug screening and it is a further object of the present invention to provide a method for identifying compounds which modulate the activity of the gene product, said method comprising exposing cells containing heterologous DNA encoding gene product, wherein said cells produce functional gene product, to at least one compound or mixture of compounds or signal whose ability to  
20 modulate the activity of said gene product is sought to be determined, and thereafter monitoring said cells for changes caused by said modulation. Such an assay enables the identification of modulators, such as agonists, antagonists and allosteric modulators, of the gene product. As used herein, a compound or signal that modulates the activity of gene product refers to a compound that alters the activity of gene product in such a way that the  
25 activity of gene product is different in the presence of the compound or signal (as compared to the absence of said compound or signal).

Cell-based screening assays can be designed by constructing cell lines in which the expression of a reporter protein, i.e. an easily assayable protein, such as  $\beta$  galactosidase, chloramphenicol acetyltransferase (CAT) or luciferase, is dependent on gene product. Such an

assay enables the detection of compounds that directly modulate gene product function, such as compounds that antagonise gene product, or compounds that inhibit or potentiate other cellular functions required for the activity of gene product.

5 The present invention also provides a method to exogenously affect gene product dependent processes occurring in cells. Recombinant gene product producing host cells, e.g. mammalian cells, can be contacted with a test compound, and the modulating effect(s) thereof can then be evaluated by comparing the gene product-mediated response in the presence and absence of test compound, or relating the gene product-mediated response of test cells, or control cells (i.e., cells that do not express gene product), to the presence of the compound.

## 10 NUCLEIC ACID LIBRARIES

The method of the present invention is useful for sorting libraries of nucleic acids. Herein, the terms "library", "repertoire" and "pool" are used according to their ordinary signification in the art, such that a library of nucleic acids encodes a repertoire of gene products. In general, libraries are constructed from pools of nucleic acids and have properties,  
15 which facilitate sorting. Initial selection of a nucleic acid from a library of nucleic acids using the present invention will in most cases require the screening of a large number of variant nucleic acids. Libraries of nucleic acids can be created in a variety of different ways, including the following.

Pools of naturally occurring nucleic acids can be cloned from genomic DNA or cDNA  
20 (Sambrook *et al.*, 1989 *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.) ; for example, phage antibody libraries, made by PCR amplification repertoires of antibody genes from immunised or unimmunised donors have proved very effective sources of functional antibody fragments (Winter *et al.*, 1994 *Annu Rev Immunol*, 12, 433-55.; Hoogenboom, H. R. (1997) *Trends Biotechnol.*, 15, 62-70). Designing  
25 and optimizing library selection strategies for generating high-affinity antibodies. *Trends Biotechnol.* 15, 62-70; Hoogenboom, H.R. (1997) *Trends Biotechnol.*, 15, 62-70). Libraries of genes can also be made by encoding all (see for example Smith, G.P. (1985) *Science*, 228, 1315-7; Parmley, S.F. and Smith, G.P. (1988) *Gene*, 73, 305-18) or part of genes (see for

example Lowman *et al.*, (1991) *Biochemistry*, 30, 10832-8) or pools of genes (see for example Nissim, A., Hoogenboom *et al.*, (1994) *Embo J*, 13, 692-8) by a randomised or doped synthetic oligonucleotide. Libraries can also be made by introducing mutations into a nucleic acids or pool of nucleic acids 'randomly' by a variety of techniques *in vivo*, including; using

5 'mutator strains', of bacteria such as *E. coli mutD5* (Liao *et al.*, (1986) *Proc Natl Acad Sci U S A*, 83, 576-80; Yamagishi *et al.*, (1990) *Protein Eng*, 3, 713-9; Low *et al.*, (1996), *J Mol Biol*, 260, 359-68); using the antibody hypermutation system of B-lymphocytes (Yelamos *et al.*, (1995), *Nature*, 376, 225-9). Random mutations can also be introduced both *in vivo* and *in vitro* by chemical mutagens, and ionising or UV irradiation (see Friedberg *et al.*, 1995, *DNA*

10 *repair and mutagenesis*. ASM Press, Washington D.C), or incorporation of mutagenic base analogues (Freese, 1959, *J. Mol. Biol.*, 1, 87; Zaccolo *et al.*, (1996), *J Mol Biol*, 255, 589-603). 'Random' mutations can also be introduced into genes *in vitro* during polymerisation for example by using error-prone polymerases (Leung *et al.*, (1989), *Technique*, 1, 11-15).

Further diversification can be introduced by using homologous recombination either *in vivo* (Kowalczykowski *et al.*, (1994) *Microbiol Rev*, 58, 401-65 or *in vitro* (Stemmer, (1994), *Nature*, 370, 389-9.; Stemmer, (1994) *Proc Natl Acad Sci U S A*, 91, 10747-51).

15

#### AGENT

As used herein, the term "agent" includes but is not limited to an atom or molecule, wherein a molecule may be inorganic or organic, a biological effector molecule and/or a

20 nucleic acid encoding an agent such as a biological effector molecule, a protein, a polypeptide, a peptide, a nucleic acid, a peptide nucleic acid (PNA), a virus, a virus-like particle, a nucleotide, a ribonucleotide, a synthetic analogue of a nucleotide, a synthetic analogue of a ribonucleotide, a modified nucleotide, a modified ribonucleotide, an amino acid, an amino acid analogue, a modified amino acid, a modified amino acid analogue, a steroid, a

25 proteoglycan, a lipid, a fatty acid and a carbohydrate. An agent may be in solution or in suspension (e.g., in crystalline, colloidal or other particulate form). The agent may be in the form of a monomer, dimer, oligomer, etc, or otherwise in a complex.

## POLYPEPTIDE

As used herein, the terms "peptide", "polypeptide" and "protein" refer to a polymer in which the monomers are amino acids and are joined together through peptide or disulfide bonds. "Polypeptide" refers to either a full-length naturally-occurring amino acid chain or a  
5 "fragment thereof" or "peptide", such as a selected region of the polypeptide that binds to another protein, peptide or polypeptide in a manner modulatable by a ligand, or to an amino acid polymer, or a fragment or peptide thereof, which is partially or wholly non-natural. "Fragment thereof" thus refers to an amino acid sequence that is a portion of a full-length polypeptide, between about 8 and about 500 amino acids in length, preferably about 8 to about  
10 300, more preferably about 8 to about 200 amino acids, and even more preferably about 10 to about 50 or 100 amino acids in length. "Peptide" refers to a short amino acid sequence that is 10-40 amino acids long, preferably 10-35 amino acids. Additionally, unnatural amino acids, for example,  $\beta$ -alanine, phenyl glycine and homoarginine may be included. Commonly encountered amino acids, which are not gene-encoded, may also be used in the present  
15 invention. All of the amino acids used in the present invention may be either the D- or L-optical isomer. The L-isomers are preferred. In addition, other peptidomimetics are also useful, e.g. in linker sequences of polypeptides of the present invention (see Spatola, (1983), *in Chemistry and Biochemistry of Amino Acids, Peptides and Proteins*, Weinstein, ed., Marcel Dekker, New York, p. 267). A "polypeptide binding molecule" is a molecule, preferably a  
20 polypeptide, protein or peptide, which has the ability to bind to another polypeptide, protein or peptide. Preferably, this binding ability is modulatable by a ligand.

The term "synthetic", as used herein, means that the process or substance described does not ordinarily occur in nature. Preferably, a synthetic substance is defined as a substance which is produced by *in vitro* synthesis or manipulation.

25 The term 'molecule' is used herein to refer to any atom, ion, molecule, macromolecule (for example polypeptide), or combination of such entities. The term 'ligand' may be used interchangeably with the term 'molecule'. Molecules according to the invention may be free in solution, or may be partially or fully immobilised. They may be present as discrete entities, or may be complexed with other molecules. Preferably, molecules according to the invention

include polypeptides displayed on the surface of bacteriophage particles. More preferably, molecules according to the invention include libraries of polypeptides presented as integral parts of the envelope proteins on the outer surface of bacteriophage particles. Methods for the production of libraries encoding randomised polypeptides are known in the art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acids, and not for stop codons.

### EXAMPLES

#### Example 1. Construction of *Taq* polymerase expression plasmids

- 10       The *Taq* polymerase open reading frame is amplified by PCR from *Thermus aquaticus* genomic DNA using primers 1 & 2, cut with *Xba*I & *Sal*I and ligated into pASK75 (Skerra A. 1994, *Gene* 151, 131)) cut with *Xba*I & *Sal*I. pASK75 is an expression vector which directs the synthesis of foreign proteins in *E. coli* under transcriptional control of the tetA promoter /operator.
- 15       Clones are screened for inserts using primers 3, 4 and assayed for expression of active *Taq* polymerase (*Taq* pol) (see below). The inactive *Taq* pol mutant D785H/E786V is constructed using Quickchange mutagenesis (Stratagene). The mutated residues are critical for activity (Doublié S. *et al*, 1998, *Nature* 391, 251; Kiefer J.R. *et al*, 1998, *Nature* 391,304). Resulting clones are screened for mutation using PCR screening with primers 3, 5 and diagnostic
- 20       digestion of the products with PmlI. Mutant clones are assayed for expression of active *Taq* pol (see below).

#### Example 2. Protein Expression and Activity Assay

- Transformed TG1 cells are grown in 2xTY 0.1mg/ml ampicillin. For expression, overnight cultures are diluted 1/100 into fresh 2xTY medium and grown to OD<sub>600</sub>=0.5 at 37 °C. Protein expression is induced by addition of anhydro tetracycline to a final concentration
- 25       of 0.2 µg/ml. After 4 hours further incubation at 37 °C, cells are spun down, washed once, and

re-suspended in an equal volume of 1 X SuperTaq polymerase buffer ( 50mM KCl, 10mM Tris-HCl (pH9.0), 0.1% TritonX-100, 1.5mM MgCl<sub>2</sub>) (HT Biotechnology Ltd, Cambridge UK).

- Washed cells are added directly to a PCR reaction mix (2 $\mu$ l per 30 $\mu$ l reaction volume) comprising template plasmid (20ng), primers 4 and 5 (1  $\mu$ M each), dNTPs (0.25mM), 1 X SuperTaq polymerase buffer, and overlaid with mineral oil. Reactions are incubated for 10 min at 94 °C to release *Taq* pol from the cells and then thermocycled with 30 cycles of the profile 94 °C (1 min), 55 °C (1 min), 72 °C (2min).

### Example 3. Emulsification of Amplification Reactions

- Emulsification of reactions is carried out as follows. 200 $\mu$ l of PCR reaction mix (*Taq* expression plasmid (200ng), primers 3 and 4 (1 $\mu$ M each ), dNTPs (0.25mM), *Taq* polymerase (10 units)) is added dropwise (12 drops/min) to the oil phase (mineral oil (Sigma)) in the presence of 4.5% (v/v) Span 80 (Fluka), 0.4% (v/v) Tween 80 (Sigma) and 0.05% (v/v) Triton X100 (Sigma) under constant stirring (1000rpm) in 2ml round bottom biofreeze vials (Costar, Cambridge MA). After complete addition of the aqueous phase, stirring is continued for a further 4 minutes. Emulsified mixtures are then transferred to 0.5 ml thin-walled PCR tubes (100 $\mu$ l/tube) and PCR carried out using 25 cycles of the profile 94 °C (1 min), 60 °C (1 min), 72 °C (3min) after an initial 5 min incubation at 94 °C. Reaction mixtures are recovered by the addition of a double volume of ether, vortexing and centrifugation for 2 minutes prior to removal of the ether phase. Amplified product is visualised on by gel electrophoresis on agarose gels using standard methods (see for example J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning: A Laboratory Manual, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press).



For emulsification of whole cells expressing *Taq* polymerase, the protocol is modified in the following way: *Taq* expression plasmid and *Taq* polymerase in the reaction cocktail are omitted and instead  $5 \times 10^8$  induced *E.coli* TG1 cells (harbouring the expressed *Taq* polymerase as well as the expression plasmid) are added together with the additive  
5 tetramethyl ammonium chloride (50  $\mu$ M), and RNase (0.05% w/v, Roche, UK). The number of PCR cycles is also reduced to 20.

#### Example 4. Self-Replication of the Full-Length wt *Taq* gene

In order to test genotype-phenotype linkage during self-replication, we mixed cells expressing either wild type *Taq* polymerase (wt *Taq*) or the poorly active (under the buffer  
10 conditions) Stoffel fragment (sf *Taq*) (F. C. Lawyer, *et al.*, *PCR Methods Appl* 2, 275-87 (1993)) at a 1:1 ratio and subjected them to CSR either in solution or in emulsion. In solution the smaller sf *Taq* is amplified preferentially. However, in emulsion there is almost exclusive self-replication of the full-length wt *Taq* gene (Figure 3B). The number of bacterial cells is adjusted such that the majority of emulsion compartments contain only a single cell. However,  
15 because cells are distributed randomly among compartments, it is unavoidable that a minor fraction will contain two or more cells. As compartments do not appear to exchange template DNA (Figure 3A), the small amount of sf *Taq* amplification in emulsion is likely to originate from these compartments. Clearly, their abundance is low and, as such, unlikely to affect selections. Indeed, in a test selection, a single round of CSR is sufficient to isolate wt *Taq*  
20 clones from a  $10^6$ -fold excess of an inactive *Taq* mutant.

Using error-prone PCR, we prepared two repertoires of random *Taq* mutants (L1 (J. P. Vartanian, M. Henry, S. Wain-Hobson, *Nucleic Acid Res.* 24, 2627-2631 (1996)) and L2 (M. Zaccolo, E. Gherardi, *J Mol Biol* 285, 775-83 (1999).) Only 1-5% of L1 or L2 clones are active, as judged by PCR, but a single round of CSR selection for polymerase activity under  
25 standard PCR conditions increased the proportion of active clones to 81% (L1\*) and 77% (L2\*).

### Example 5. Mutagenic PCR

*Taq* polymerase gene variants are constructed using two different methods of error-prone PCR.

- 5       The first utilises the nucleoside analogues dPTP and dLTP (Zaccolo *et al*, (1996) *J Mol Biol.* 255, 589-603). Briefly, a 3-cycle PCR reaction comprising 50mM KCl, 10mM Tris-HCl (pH9.0), 0.1% TritonX-100, 2 mM MgCl<sub>2</sub>, dNTPS (500µM), dPTP (500µM), dLTP (500µM), 1 pM template DNA, primers 8 and 9 (1 µM each), *Taq* polymerase (2.5 units) in a total volume of 50µl is carried out with the thermal profile 94 °C (1 min.), 55 °C (1 min.), 72 °C (5 min). A 2µl aliquot is then transferred to a 100 µl standard PCR reaction comprising
- 10       50mM KCl, 10mM Tris-HCl (pH9.0), 0.1% TritonX-100, 1.5 mM MgCl<sub>2</sub>, dNTPS (250µM), primers 6 and 7 (1 µM each), *Taq* polymerase (2.5 units). This reaction is cycled 30 x with the profile 94 °C (30 seconds), 55 °C (30 seconds), 72 °C (4 minutes). Amplified product is gel-purified, and cloned into pASK75 as above to create library L2.
- 15       The second method utilises a combination of biased dNTPs and MnCl<sub>2</sub> to introduce errors during PCR. The reaction mix comprises 50mM KCl, 10mM Tris-HCl (pH9.0), 0.1% TritonX-100, 2.5 mM MgCl<sub>2</sub>, 0.3 mM MnCl<sub>2</sub>, 1 pM template DNA, dTTP, dCTP, dGTP (all 1mM), dATP (100µM) primers 8 and 9 (1 µM each) and *Taq* polymerase (2.5 units). This reaction is cycled 30 x with the profile 94 °C (30 seconds), 55 °C (30 seconds), 72 °C (4
- 20       minutes), and amplified products cloned as above to create library L1.

### Example 6. Selection Protocol

- For selection of active polymerases, PCR reactions within emulsions are carried out as described above but using primers 8, 9. For selection of variants with increased thermostability, emulsions are preincubated at 99 °C for up to 7 minutes prior to cycling as
- 25       above. For selection of variants with increased activity in the presence of the inhibitor heparin, the latter is added to concentrations of 0.08 and 0.16 units/µl and cycling carried out as above. Detailed protocols are set out in further Examples below.

Amplification products resulting from compartments containing an active polymerase are extracted from emulsion with ether as before and then purified by standard phenol-chloroform extraction. 0.5 volumes of PEG/MgCl<sub>2</sub> solution (30% v/v PEG 800, 30mM MgCl<sub>2</sub>) is next added, and after mixing centrifugation carried out at 13,000 RPM for 10 minutes at room temperature. The supernatant (containing unincorporated primers and dNTPs) is discarded and the pellet re-suspended in TE. Amplified products are then further purified on spin-columns (Qiagen) to ensure complete removal of primers. These products are then re-amplified using primers 6, 7 (which are externally nested to primers 8 and 9) in a standard PCR reaction, with the exception that only 20 cycles are used. Re-amplified products are gel-purified and re-cloned into pASK75 as above. Transformants are plated and colonies screened as below. The remainder are scraped into 2xTY/0.1mg/ml ampicillin, diluted down to OD<sub>600</sub>=0.1 and grown/induced as above for repetition of the selection protocol.

#### Example 7. Colony Screening Protocol

Colonies are picked into a 96 well culture dish (Costar), grown and induced for expression as above. For screening, 2µl of cells are used in a 30µl PCR reaction to test for activity as above in a 96 well PCR plate (Costar) using primers 4 and 5. A temperature gradient block is used for the screening of selectants with increased thermostability. Reactions are preincubated for 5 minutes at temperatures ranging from 94.5 to 99°C prior to standard cycling as above with primers 4 and 5 or 3 and 4. For screening of heparin-compatible polymerases, heparin is added to 0.1 units/30µl during the 96-well format colony PCR screen. Active polymerases are then assayed in a range of heparin concentrations ranging from 0.007 to 3.75 units/30µl and compared to wildtype.

#### Example 8. Assay for Catalytic Activity of Polymerases

K<sub>cat</sub> and K<sub>m</sub> (dTTP) are determined using a homopolymeric substrate (Polesky *et al.*, (1990) *J. Biol. Chem.* 265:14579-91). The final reaction mix (25µl) comprises 1X SuperTaq buffer (HT Biotech), poly(dA).oligo(dT)(500nM, Pharmacia), and variable concentrations of [ $\alpha$ -<sup>32</sup>P]dTTP (approx. 0.01 Ci/mMole). The reaction is initiated by addition of 5µl enzyme in

1X SuperTaq buffer to give a final enzyme concentrations between 1-5nM. Reactions are incubated for 4 minutes at 72°C, quenched with EDTA as in example 14, and applied to 24mm DE-81 filters. Filters are washed and activity measured as in example 14. Kinetic parameters are determined using the standard Lineweaver-Burke plot. Experiments using 50%  
5 reduced homopolymer substrate show no gross difference in incorporation of dTTP by polymerase, indicating it is present in sufficient excess to validate the kinetic analysis protocol used.

#### Example 9. Standard PCR in Aqueous Compartments Within an Emulsion

To establish whether conditions in the aqueous compartments present in an emulsion  
10 are permissive for catalysis, a standard reaction mix is emulsified and PCR carried out. This leads to amplification of the correct sized *Taq* polymerase gene present in the plasmid template, with yields sufficient yields to allow visualisation using standard agarose gel electrophoresis.

#### Example 10. Emulsification of *E. coli* expressing *Taq* Polymerase and Subsequent PCR to 15 Amplify Polymerase Gene

*E. coli* cells expressing *Taq* polymerase are emulsified and PCR carried out using primers flanking the polymerase cassette in the expression vector. Emulsification of up to  $5 \times 10^8$  cells (per 600µl total volume) leads to discernible product formation as judged by agarose gel electrophoresis. The cells therefore segregate into the aqueous compartments where  
20 conditions are suitable for self-amplification of the polymerase gene by the expressed *Taq* polymerase. Similar emulsions are estimated to contain about  $1 \times 10^{10}$  compartments per ml (Tawfik D. & Griffiths A.D. (1998) *Nature Biotech.* 16, 652). The large number of cells that can be emulsified allows for selection from diverse repertoires of randomised protein.

#### Example 11. Maintenance of Genotype-Phenotype Linkage in Emulsion

25 To be viable for a selection method, the majority of aqueous compartments in the emulsion should harbour a single cell, and the integrity of compartments should be maintained

during thermal cycling. This is tested by including in the emulsion cells harbouring a competitor template distinguishable by its smaller size.

*E. coli* expressing *Taq* polymerase are co-emulsified with *E. coli* expressing the Stoffel fragment at a ratio of one to one. The Stoffel fragment is poorly active under the conditions used in emulsion, and thus amplification of its expression cassette by the same primer pair used for *Taq* self-amplification is the result of co-compartmentalisation with a cell expressing active *Taq* polymerase or leakage of *Taq* polymerase between compartments. After PCR, the vast majority of products are found to correspond to the active *Taq* polymerase gene thus validating the premise of one cell per durable compartment (see Fig. 2, Ghadessy *et al* (2001), *PNAS*, 98, 4552).

#### Example 12. Test Selection of Active over Inactive *Taq* polymerase

To demonstrate that the method can select for potentially rare variants, a  $10^6$  fold excess of cells expressing inactive polymerase over those expressing the active form are co-emulsified. After PCR and cloning of amplified product, a single expression screen using a 96 well format indicated a  $10^4$  fold enrichment for the active polymerase.

#### Example 13. Directed Evolution of *Taq* Polymerase Variants with Increased Thermal Stability

Polymerases with increased thermostability are of potential practical importance, reducing activity loss during thermocycling and allowing higher denaturation temperatures for the amplification of GC rich templates. Thus, we first used the selection method of our invention for the directed evolution of *Taq* variants with increased thermostability, starting from preselected libraries (L1\*, L2\*) and progressively increasing the temperature and duration of the initial thermal denaturation. After 3 rounds of selection, we isolated T8 (Table 1), a *Taq* clone with an 11-fold longer half-life at 97.5°C than the already thermostable wt *Taq* enzyme (Table 2), making T8 the most thermostable member of the Pol I family on record (Clones are screened and marked by a PCR assay. Briefly, 2µl of induced cells are added to 30µl PCR mix and amplification of a 0.4kb fragment is assayed under selection conditions (e.g. increasing amounts of heparin). Thermostability and heparin resistance of purified His

tagged wt and mutant *Taq* clones is determined as in (Lawyer et al., PCR Methods Appl 2, 275-287 (1993); Lawer et al., J Biol Chem 264, 6427-37 (1989) using activated salmon sperm DNA and normalized enzyme concentrations). Mutations conferring thermostability to T8 (and to a majority of less thermostable mutants) cluster in the 5'-3' exonuclease domain (Table 1). Indeed, truncation variants of *Taq* polymerase (F. C. Lawyer, et al., (1993) *PCR Methods Appl* 2, 275-87; W. M. Barnes, (1992) *Gene* 112, 29-35) lacking the exonuclease domain show improved thermostability, suggesting it may be less thermostable than the main polymerase domain. The lower thermostability of the exonuclease domain may have functional significance (for example reflecting a need for greater flexibility), as the stabilizing mutations in T8 appear to reduce exonuclease activity (approx. 5-fold) (5'-3' exonuclease activity is determined essentially as in (Y. Xu, et al., *J Mol Biol* 268, 284-302 (1997)) but in 1x*Taq* buffer with 0.25mM dNTP's and the 22-mer oligonucleotide of (Y. Xu, et al., *J Mol Biol* 268, 284-302 (1997)) 5' labelled with Cy5 (Amersham). Steady-state kinetics are measured as in (A. H. Polesky, T. A. Steitz, N. D. Grindley, C. M. Joyce, *J Biol Chem* 265, 14579-91 (1990) using the homopolymeric substrate poly(dA)<sub>200</sub> (Pharmacia) and oligo(dT)<sub>40</sub> primer at 50°C.) (at least at low temperature).

Round	<i>Taq</i> variant	Thermo-stability*	Heparin Resistance*
	<i>Taq</i> <sub>wt</sub>	1	1
1	T646 (G46V, A109P, F285L)	2x	n.d.
	T788 ( <u>F73S</u> , <u>R205K</u> , <u>K219E</u> , <u>M236T</u> , <u>A608V</u> )	4x	n.d.
2	T9 (F278L, P298S)	4x	n.d.
	T13 ( <u>R205K</u> , <u>K219E</u> , <u>M236T</u> , <u>A608V</u> )	7x	n.d.
3	T8 ( <u>F73S</u> , <u>R205K</u> , <u>K219E</u> , <u>M236T</u> , <u>E434D</u> , <u>A608V</u> )	11x	<0.5x
1	H32 (E9K, P93S, K340E, Q534R, T539A, V703A, R778K)	n.d.	8x
2	H94 ( <u>K225E</u> , L294P, A454S, L461R, <u>D578G</u> , <u>N583S</u> )	n.d.	32x
3	H15 ( <u>K225E</u> , E388V, K540R, <u>D578G</u> , <u>N583S</u> , M747R)	0.3x	130x

\* as judged by PCR (relative to *Taq<sub>wt</sub>*), at 97.5°C

\*\* as judged by PCR (relative to *Taq<sub>wt</sub>*)

Table 1 : Properties of selected clones. Clones in bold are related through underlined mutations. Clones are ranked in relation to wt *Taq*.

Two libraries of *Taq* polymerase variants generated using error-prone PCR are expressed in *E. coli* (library L1, 8x10<sup>7</sup> clones, library L2 2x10<sup>7</sup> clones; see example 5) and emulsified as before. The first round of PCR is carried out to enrich for active variants using the standard *Taq* polymerase thermocycling profile outlined above. Enriched amplification products are purified, and recloned to generate libraries comprising of active variants (L1\*, L2\*; approx 10<sup>6</sup> clones for each library). A screen of the L1\* and L2\* libraries respectively showed 81% and 77% of randomly picked clones to be active.

Selective pressure is applied to the L1\* and L2\* libraries during the next round of PCR by pre-incubating emulsions at 99°C for 6 or 7 minutes prior to the normal PCR cycle. Under these conditions, the wild type *Taq* polymerase loses all activity. Amplified products are enriched and cloned as above and a 96-well expression screen used to select for active variants under normal PCR conditions. This yielded 7 clones from the L2\* library and 10 clones from the L1\* library. These are then screened for increased thermostability using a temperature gradient PCR block, with a 5 minute pre-incubation at temperatures of 94.5 to 99°C prior to standard cycling. As judged by gel electrophoresis, 5 clones from each library are present with increased thermostability compared to wild type. These mutants are able to efficiently amplify the 320 b.p. target after pre-incubation at 99°C for 5 minutes. The wild type enzyme has no discernible activity after pre-incubation at temperatures above 97°C for 5 minutes or longer.

Example 14. Assay for Thermal Stability of Polymerase

Thermal inactivation assays of WT and purified His-tagged polymerases are carried out in a standard 50  $\mu$ l PCR mixture comprising 1X Super*Taq* buffer (HT Biotech), 0.5ng plasmid DNA template, 200 $\mu$ M each of dATP, dTTP, and dGTP, primers 3 and 4 (10 $\mu$ M), and polymerase (approximately 5nM). Reaction mixtures are overlaid with oil and incubated at 97.5°C, with 5 $\mu$ l aliquots being removed and stored on ice after defined intervals. These aliquots are assayed in a 50 $\mu$ l activity reaction buffer comprising 25mM N-tris[hydroxymethyl-3-amino-propanesulfonic acid (TAPS)(pH9.5), 1 mM  $\beta$ -mercaptoethanol, 2mM MgCl<sub>2</sub>, 200 $\mu$ M each dATP, dTTP, and dGTP, 100 $\mu$ M [ $\alpha$ -<sup>32</sup>P]dCTP (0.05 Ci/mmmole), and 250  $\mu$ g/ml activated salmon sperm DNA template. Reactions are incubated for 10 minutes at 72°C, stopped by addition of EDTA (25mM final). Reaction volumes are made up to 500 $\mu$ l with solution S (2mM EDTA, 50ug/ml sheared salmon sperm DNA) and 500 $\mu$ l 20% TCA (v/v) / 2% sodium pyrophosphate (v/v) added. After 20 minutes incubation on ice, reactions are applied to 24mm GF/C filters (Whatman). Unincorporated nucleotides are removed by 3 washes with 5% TCA(v/v), 2% sodium pyrophosphate (v/v) followed by two washes with 96% ethanol (v/v). Dried filters are counted in scintillation vials containing Ecoscint A (National Diagnostics). The assay is calibrated using a known amount of the labeled dCTP solution (omitting the washes).

Example 15. Directed Evolution of *Taq* Polymerase Variants with Increased Activity in the Presence of the Inhibitor Heparin

As indicated above, the methods of our invention can also be used to evolve resistance to an inhibitor of enzymatic activity. Heparin is a widely used anticoagulant, but also a potent inhibitor of polymerase activity, creating difficulties for PCR amplifications from clinical blood samples (J. Satsangi, D. P. Jewell, K. Welsh, M. Bunce, J. I. Bell, *Lancet* 343, 1509-10 (1994)). While heparin can be removed from blood samples by various procedures, these can be both costly and time-consuming. The availability of a heparin-compatible polymerase would therefore greatly improve characterisation of therapeutically significant amplicons, and



obviate the need for possibly cost-prohibitive heparinase treatment of samples (Taylor A.C. (1997) *Mol. Ecol* 6, 383).

The L1\* and L2\* libraries are combined, and selected in emulsion for polymerases active in up to 0.16 units heparin per  $\mu\text{l}$ . After a single round, 5 active clones are isolated in the 96 well PCR screen incorporating 0.1 units/30 $\mu\text{l}$  reaction, with the wild type showing no activity. Titration shows that 4 of these clones to be active in up to four times the amount of heparin inhibiting wild type (0.06units/30 $\mu\text{l}$  versus 0.015units/30 $\mu\text{l}$ ). The other clone is active in up to eight times the amount of heparin inhibiting wild type (0.12units/30 $\mu\text{l}$  versus 0.015units/30 $\mu\text{l}$ ).

Using selection in the presence of increasing amounts of heparin, we isolated H15, a *Taq* variant functional in PCR at up to 130-times the inhibitory concentration of heparin (Table 2). Intriguingly, heparin resistance conferring mutations also cluster, in this case in the base of the finger and thumb polymerase subdomains, regions involved in binding duplex DNA. Indeed, judging from a recent high-resolution structure of a *Taq*-DNA complex (Y. Li, S. Korolev, G. Waksman, *EMBO J* 17, 7514-25 (1998)) four out of six residues mutated in H15 (K540, D578, N583, M747) directly contact either template or product strand (as shown in Figure 7). H15 mutations appear to be neutral (or mutually compensating) as far as affinity for duplex DNA is concerned (while presumably reducing affinity for heparin) (Table 2) ( $K_D$  for DNA is determined using BIAcore. Briefly, the 68-mer used in (M. Astatke, N. D. Grindley, C. M. Joyce, *J Biol Chem* 270, 1945-54 (1995)) is biotinylated at the 5' end and bound to a SA sensorchip and binding of polymerases is measured in 1x *Taq* buffer (see above) at 20°C. Relative  $K_D$  values are estimated by the PCR ranking assay using decreasing amounts of template). The precise molecular basis of heparin inhibition is not known, but our results strongly suggest overlapping (and presumably mutually exclusive) binding sites for DNA and heparin in the polymerase active site, lending support to the notion that heparin exerts its inhibitory effect by mimicking and competing with duplex DNA for binding to the active site. Our observation that heparin inhibition is markedly reduced under conditions of excess template DNA, (see (Clones are screened and ranked by a PCR assay. Briefly, 2 $\mu\text{l}$  of induced cells are added to 30 $\mu\text{l}$  PCR mix and amplification of a 0.4kb fragment is assayed under selection conditions (e.g. increasing amounts of heparin). Thermostability and heparin resistance of

purified His tagged wt and mutant *Taq* clones is determined as in (F. C. Lawyer, et al., *PCR Methods Appl* 2, 275-87 (1993); F. C. Lawyer, et al., *J Biol Chem* 264, 6427-37 (1989)) using activated salmon sperm DNA and normalized enzyme concentrations, Table 2) appears consistent with this hypothesis.

5

Table 2: Properties of selected *Taq* clones

<i>Taq</i> clone	$T_{1/2}(97.5^{\circ}\text{C})$ (min)	Heparin resistance (units/ml)		$K_D$ ( $\text{nM}^{-1}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_M$ $\sigma\text{TTP}$ ( $\mu\text{M}$ )	5'-3' exo activity	Mutation Rate <sup>§</sup>
<i>Taq</i> <sup>*</sup>	n.d.	n.d.	0.6 <sup>***</sup>	0.8 <sup>†</sup>	4.0 <sup>‡</sup>	43.2	n.d.	1.1
<i>Taq</i> <sub>wt</sub>	1.5 <sup>**</sup>	90 <sup>**</sup>	0.6 <sup>***</sup>	0.8	9.0	45.0	1	1
T8	16.5 <sup>**</sup>	n.d.	0.3 <sup>***</sup>	1.2	8.8	48.6	0.2	1.2
H15	0.3 <sup>**</sup>	1750 <sup>*</sup>	84 <sup>***</sup>	0.79	6.8	47.2	1.5	0.9

\* commercial *Taq* preparation (HT Biotechnology), \*\* with N-terminal His<sub>6</sub> tag, measured by CTP<sup>32</sup> incorporation into salmon sperm DNA, \*\*\* no tag, measured by PCR assay, <sup>†</sup> *Taq*, published value:  $1\text{nM}^{-1}$  (J), Klenow (Cambio),  $4\text{nM}^{-1}$ , <sup>‡</sup> *E.coli* DNA Pol I, published value:  $3.8\text{ s}^{-1}$  (A. H. Polesky, T. A. Steitz, N. D. Grindley, C. M. Joyce, *J Biol Chem* 265, 14579-91 (1990)), <sup>§</sup> in relation to *Taq*<sub>wt</sub> measured by mutS ELISA (Genecheck) (P. Debbie, et al., *Nucleic Acids Res* 25, 4825-4829 (1997).), Pfu (Stratagene): 0.2.

10

15

#### Example 16: Template evolution in emulsion selection

A classic outcome of in vitro replication experiments is an adaptation of the template sequence towards more rapid replication (S. Spiegelman, *Q. Rev. Biophys.* 4, 213-253 (1971)). Indeed, we also observe template evolution through silent mutations. Unlike the coding mutations (AT to GC vs. GC to AT / 29 vs. 16), non-coding mutations display a striking bias (AT to GC vs. GC to AT / 0 vs. 42) towards decreased GC content, generally thought to promote more efficient

20

replication by facilitating strand separation and destabilizing secondary structures. Apart from selecting for adaptation, our method may also select for adaptability; i.e. polymerases might evolve towards an optimal, presumably higher, rate of self-mutation (M. Eigen, *Naturwissenschaften* 58, 465-523 (1971)). Indeed, mutators can arise spontaneously in asexual bacterial populations under adaptive stress (F. Taddei, et al., *Nature* 387, 700-2 (1997); P. D. Sniegowski, P. J. Gerrish, R. E. Lenski, *Nature* 387, 703-5 (1997)). By analogy, it could be argued that our method might favour polymerase variants that are more error-prone and hence capable of faster adaptive evolution. However, none of the selected polymerases displayed increased error rates (Table 2). Eliminating recombination and decreasing the mutational load during our method cycle may increase selective pressures towards more error-prone enzymes.

#### Example 17 Assay for Heparin Tolerance of Polymerases

Heparin tolerance of polymerases is assayed using a similar assay to that for thermal stability. Heparin is serially diluted into the activity buffer (0-320 units/45µl) and 5µl of enzyme in the standard PCR mixture above are added. Reactions are incubated and incorporation assayed as above.

#### Example 18. Selection for *Taq* Variants with Increased Ability to Extend from a 3' Mismatched Base

The primers used are Primer 9 (LMB388ba5WA) and Primer 10 (8fo2WC). This primer combination presents polymerase variants with a 3' purine-purine mismatch (A-G), and a 3' pyrimidine-pyrimidine mismatch (C-C). These are the mismatches least tolerated by *Taq* polymerase (Huang et al., 1992, *Nucleic Acids Res* 20(17):4567-73) and are poorly extended.

The selection protocol is essentially the same as before, except that these two primers are used in emulsion. Extension time is also increased to 8 minutes. After two rounds of selection, 7 clones are isolated which display up to a 16-fold increase in extension off the mismatch as judged by a PCR ranking assay (see example 2: using primers 5 and 11) and standardised for activity using the normal primer pair. These clones are subsequently shuffled

back into the original L1\* and L2\* libraries along with wild type *Taq* and the selection process repeated, albeit with a lower number of cycles (10) during the CSR reaction. This round of selection yielded numerous clones, the best of which displayed up to 32-fold increase in mismatch extension as judged by PCR (see example 2) using primers 5 and 11.

- 5 Incorporation of an incorrect base pair by *Taq* polymerase can stall the polymerisation process as certain mismatches (see above) are poorly extended by *Taq*. As such, *Taq* polymerase alone cannot be used in the amplification of large (>6Kb) templates (Barnes). This problem can be overcome by supplementing *Taq* with a polymerase that has a 3'-5' exonuclease activity (eg Pfu polymerase) that removes incorrectly incorporated bases and allows resumption of
- 10 polymerisation by *Taq*. The clones above are therefore investigated for their ability to carry out amplification of large DNA fragments (long-distance PCR) from a lambda DNA template, as incorporation of an incorrect base would not be expected to stall polymerisation. Using primers 12 (LBA23) and 13 (LFO46) (1uM each) in a 50ul PCR reaction containing 3ng lambda DNA (New England Biolabs) dNTPs (0.2 mM), 1x PCR buffer (HT Biotech) clone
- 15 M1 is able to amplify a 23Kb fragment using 20 repetitions of a 2-step amplification cycle (94 °C, 15 seconds; 68 °C, 25 minutes). Wild type polymerase is unable to extend products above 13 Kb using the same reaction buffer. Commercial *Taq* (Perkin Elmer) could not extend beyond 6 Kb using buffer supplied by the manufacturer.

#### Example 19 Selection Using Self-Sustained Sequence Replication (3SR)

- 20 To demonstrate the feasibility of 3SR within emulsion, the *Taq* polymerase gene is first PCR-amplified from the parent plasmid (see example 1) using a forward primer that is designed to incorporate a T7 RNA polymerase promoter into the PCR product. A 250 µl 3SR reaction mix comprising the modified *Taq* gene (50ng), 180 units T7 RNA polymerase (USB, 63 units reverse transcriptase (HT Biotech), rNTPs (12.5mM), dNTPs (1mM), MgCl<sub>2</sub>
- 25 (10mM), primer Taqba2T7 (primer 12; 125pmoles), primer 88fo2 (primer 4; 125pmoles), 25mM Tris-HCl (pH 8.3), 50mM KCl, and 2.0mM DTT is made. 200µl of this is emulsified using the standard protocol. After prolonged incubation at room temperature, amplification of the *Taq* gene (representing a model gene size) within emulsion is seen to take place as judged by standard gel-electrophoresis.

To further expand the scope of the method, the 3SR reaction is carried out in an in-vitro transcription/translation extract (EcoPro, Novagen). The inactive taq gene (see example 1) is amplified from parental plasmid using primers 2 (TaqfoSal) and 12 (Taqba2T7). 100ng (approx.  $1 \times 10^{10}$  copies) is added to make up 100ul of the aqueous phase comprising EcoPro extract (70ul), methionine (4ul), reverse transcriptase (84 units, HT Biotech), primer 12 (Taqba2T7, 2uM), primer 13 (TaqfoLMB2, 2uM), dNTPs (250uM). The aqueous phase is emulsified into 400ul oil-phase using the standard protocol. After incubation at 37 °C overnight, the emulsion is extracted using the standard protocol and the aqueous phase further purified using a PCR-purification column (Qiagen). Complete removal of primers is ensured by treating 5ul of column eluate with 2ul ExoZap reagent (Stratagene). DNA produced in emulsion by 3SR is rescued by using 2ul of treated column eluate in an otherwise standard 50ul PCR reaction using 20 cycles of amplification and primers 6 (LMB, ref 2) and 12 (Taqba2T7). Compared to background (the control reaction where reverse transcriptase is omitted from the 3SR reaction in emulsion), a more intense correctly sized band could be seen when products are visualised using agarose gel electrophoresis. The 3SR reaction can therefore proceed in the transcription/translation extracts, allowing for the directed evolution of agents expressed in aqueous compartments.

WT *Taq* polymerase has limited reverse transcriptase activity (Perler et al., (1996) *Adv Protein Chem.* 48, 377-435). It is also known that reverse transcriptases (eg HIV reverse transcriptase that has both reverse transcriptase and polymerase activities) are considerably more error prone than other polymerases. This raises the possibility that a more error-prone polymerase (where increased tolerance for non-cognate substrate is evident) might display increased reverse transcriptase activity. The genes for *Taq* variants M1, M4 as well as the inactive mutant are amplified from parental plasmids using primers 12 (Taqba2T7) and 2 (TaqfoSal) and the 3SR reaction is carried out as above in the transcription/translation extract (Novagen) with the exception that reverse transcriptase is not exogenously added. In control reactions, methionine is omitted from the reaction mix. After 3 hours incubation at 37°C, the reaction is treated as above and PCR carried out using primer pair 6 and 12 to rescue products synthesised during the 3SR reaction. Of the clones tested, clone M4 gave a more intense correctly sized band compared to control reaction when products are visualised using agarose

gel electrophoresis. Clone M4 would therefore appear to possess some degree of reverse transcriptase activity. This result shows that it is possible to express functionally active replicases in vitro. When coupled to selection by compartmentalisation, novel replicases could be evolved.

## 5 Selection of Agents Modifying Replicase Activity

Example 19 and the following Examples describes how the methods of our invention may be employed to select an enzyme which is involved in a metabolic pathway whose final product is a substrate for the replicase. These Examples show a method for selection of nucleoside diphosphate kinase (NDP Kinase), which catalyses the transfer of a phosphate group from ATP to a deoxynucleoside diphosphate to produce a deoxynucleoside triphosphate (dNTP). Here, the selectable enzyme (NDK) provides substrates for *Taq* polymerase to amplify the gene encoding it. This selection method differs from the compartmentalized self-replication of a replicase (CSR, Ghadessy and Holliger) in that replication is a coupled process, allowing for selection of enzymes (nucleic acids and protein) that are not replicases themselves. Bacteria expressing NDK (and containing its gene on an expression vector) are co-emulsified with its substrate (in this case, dNDPs and ATP) along with the other reagents needed to facilitate its amplification (*Taq* polymerase, primers specific for the ndk gene, and buffer). Compartmentalization in a water-in-oil emulsion ensures the segregation of individual library variants. Active clones provide the dNTPs necessary for *Taq* polymerase to amplify the ndk gene. Variants with increased activity provide more substrate for its own amplification and hence post-selection copy number correlates to enzymatic activity within the constraints of polymerase activity. Additional selective pressure arises from the minimum amount of dNTPs required for polymerase activity, hence clones with increased catalytic activity are amplified preferentially at the expense of poorly active variants (selection is for kcat as well as Km).

By showing that we can evolve an enzyme whose product feeds into the polymerase reaction, we hope to eventually co-evolve multiple enzymes linked through a pathway where one enzyme's product is substrate for the next. Diversity could be introduced into two or more genes, and both genes could be co-transformed into the same expression host on

plasmids or phage. We hope to develop cooperative enzyme systems that enable selection for the synthesis of unnatural substrates and their subsequent incorporation into DNA.

#### Example 20 Induced Expression of NDP Kinase in Bacterial Cells

A pUC19 expression plasmid containing the *EcoRI/HindIII* restriction fragment with  
5 the open reading frame of Nucleoside Diphosphate Kinase from *Myxococcus Xanthus* is  
cloned. Plasmid is prepared from an overnight culture and transformed into the *ndk*<sup>-</sup>, *pykA*<sup>-</sup>,  
*pykF*<sup>-</sup> strain of *E. coli* QL1387. An overnight culture of QL1387/pUC19*ndk* is grown in the  
presence of chloramphenicol (10 µg/ml final concentration), ampicillin (100 µg/ml final  
concentration) and glucose (2%) for 14-18 hours. The overnight culture is diluted 1:100 in  
10 (2XTY, 10 µg/ml chloramphenicol, 100 µg/ml ampicillin and 0.1% glucose). Cells are grown  
to an O.D. (600 nm) of 0.4 and induced with IPTG (1mM final concentration) for 4 hours at  
37°C. After protein induction, cells are washed once in SuperTaq buffer (10 mM tris-HCL pH  
9, 50 mM KCl, 0.1% Triton X-100, 1.5 mM MgCl<sub>2</sub>, HT Biotechnology) and resuspended in  
1/10 volume of the same buffer. The number of cells is quantified by spectrophotometric  
15 analysis with the approximation of  $0.01 \times 0.600 \times 10^9 = 1 \times 10^8$  cells/ml.

#### Example 21 Phosphoryl Transfer Reaction in Aqueous Compartments Within an Emulsion

To establish whether deoxynucleoside diphosphates can be phosphorylated by NDP  
kinase in *Taq* buffer, a standard PCR reaction is carried out in which dNTPs are replaced by  
dNDPs and ATP, a donor phosphate molecule. Nucleoside diphosphate kinase is expressed  
20 from *E. coli* QL1387 (a *ndk* and pyruvate kinase deficient strain of *E. coli*) as described in the  
previous example. Cells are mixed with the PCR reaction mix.

Washed cells are added to a PCR reaction mixture (approx.  $8 \times 10^5$  cells/µl final  
concentration) containing *SuperTaq* buffer, 0.5 µM primers, 100 µM each dNDP, 400 µM  
ATP, *SuperTaq* polymerase (0.1 unit/µl final concentration, HT Biotechnology).

After breaking open the cells at 65 °C for 10 min, incubating the reaction mixture for 10 minutes at 37 °C, and thermocycling (15 cycles of 94 °C 15sec, 55 °C 30 sec, 72 °C 1min30sec), amplified products are visualized on a standard 1.5% agarose/TBE gel stained with ethidium bromide (Sambrook). The results of this experiment show that expressed NDP  
5 kinase can phosphorylate dNDPs to provide *Taq* polymerase with substrates for the PCR amplification of the *ndk* gene.

The experiment is repeated, with the additional step of emulsifying the reaction mixture with mineral oil and detergent as described above. It is found that NDP kinase is active within aqueous compartments of an emulsion

#### 10 Example 22. Compartmentalization of NDK Variants by Emulsification

The original emulsion mix allowed for the diffusion of small molecules between compartments during thermocycling. However, by adjusting the water to oil ratio and minimizing the thermocycling profile, the exchange of product and substrate between  
15 compartments is minimized, resulting in a tighter linkage of genotype to phenotype. Given the diffusion rates can be controlled by modifying the emulsion mix, it may be possible to adjust buffer conditions after emulsification, possibly allowing for greater control of selection conditions (i.e. adjusting pH with the addition of acid or base, or starting/stopping reactions with the addition of substrates or inhibitors).

20 150 µl of PCR reaction mix (*SuperTaq* buffer, 0.5 µM each primer, 100 µM each dNDP, 400 µM ATP, 0.1 unit/µl *Taq* polymerase,  $8 \times 10^5$  cells/µl of QL1387/*ndk*) are added dropwise (1 drop/5 sec) to 450 µl oil phase (mineral oil) in the presence of 4.5% v/v Span 80, 0.4% v/v Tween 80 and 0.05% v/v Triton X-100 under constant stirring in a 2 ml round bottom biofreeze vial (Corning). After addition of the aqueous phase, stirring is continued for  
25 an additional 5 minutes. Emulsion reactions are aliquoted (100 µl) into thin-walled PCR tubes and thermocycled as indicated above.



Recovery of amplified products after emulsification is carried out as follows. After thermocycling, products are recovered by extraction with 2 volumes of diethyl ether, vortexed, and centrifuged for 10 minutes in a tabletop microfuge. Amplification products are analyzed as before.

5    Example 23. Minimizing Background Kinase Activity

Background kinase activity levels are determined by emulsifying *E. coli* TG1 cells in *Taq* buffer with substrates, as described above. It is found that native nucleoside diphosphate kinase from *E. coli* retained enough activity after the initial denaturation to provide significant kinase activity in our assay. The pUC19 expression plasmid containing the *ndk* gene is  
10    transformed into a *ndk* deficient strain of *E. coli* QL1387. Compared to a catalytic knockout mutant of *mx ndk* (H117A), the background kinase activity is determined to be negligible in our assay (amplified products could not be visualized by agarose gel electrophoresis) when *ndk* is expressed from the knockout strain.

Example 24. Maintenance of the Genotype-Phenotype Linkage in Emulsion.

15    A catalytic knockout mutation (NDK H117A) of NDP kinase is co-emulsified with wild-type NDP kinase in equal amounts. The inactive mutant of *ndk* is distinguished by a smaller amplification product, since the 5' and 3' regions flanking the ORF downstream from the priming sites are removed during construction of the knockout mutant. Our emulsification procedure gives complete bias towards amplification of the active kinase, as determined by  
20    agarose gel electrophoresis.

Example 25: Method for the parallel genotyping of heterogenous populations of cells.

The approach involves compartmentation of the cells in question in the emulsion (see WO9303151) together with PCR reagents etc. and polymerase. However, instead of linking genes derived from one cell by PCR assembly, one (or several) biotinylated primers are used  
25    as well as a streptavidin coated polystyrene beads (or any other suitable means of linking primers onto beads). Thus, PCR fragments from one single cell are transferred to a single

bead. Beads are pooled, interrogated for presence of a certain mutation or allele using fluorescently labelled probes (as described for "Digital PCR") and counted by FACS. Multiplex PCR allows the simultaneous interrogation of 10 or maybe more markers. Single beads can also be sorted for sequencing.

- 5 Applications include, for example, diagnosis of asymptomatic tumors, which hinge on the detection of a very small number of mutant cells in a large excess of normal cells. The advantage of this method over cytostaining is through-put. Potentially  $10^8$ - $10^9$  cells can be interrogated simultaneously.

#### Example 25: short-patch CSR

10

- The present example relates to the selection of polymerases with low catalytic activity or processivity. Compartmentalized Self-Replication (CSR), as described, is a method of selecting polymerase variants with increased adaptation to distinct selection conditions. Mutants with increased catalytic activity have a selective advantage over ones that are less
- 15 active under the selection conditions. However, for many selection objectives (e.g. altered substrate specificity) it is likely that intermediates along the evolutionary pathway to the new phenotype will have lowered catalytic activity. For example, from kinetic studies of *E. coli* DNA polymerase I, mutations such as E710A increased affinity and incorporation of ribonucleotides at the expense of lower catalytic rates and less affinity for wild-type substrates
- 20 (deoxyribonucleotides) (F. B. Perler, S. Kumar, H. Kong, *Adv. in Prot. Chem.* 48, 377-430 (1996)). The corresponding mutant of *Taq* DNA polymerase I, E615A, could incorporate ribonucleotides into PCR products more efficiently than wild-type polymerase. However, using wild-type substrates, it is only able to synthesize short fragments and not the full-length *Taq* gene, as analyzed by agarose gel electrophoresis. Therefore it would be difficult to select
- 25 for this mutation by CSR. In another selection experiment in which Beta-glucuronidase is evolved into a  $\beta$ -galactosidase, the desired phenotype is obtained after several rounds of selection but at the expense of catalytic activity. It is also found that selected variants in the initial rounds of selection are able to catalyze the conversion of several different substrates not

utilized by either parental enzyme, and at much lower catalytic rates (T. A. Steitz, *J Biol Chem* 274, 17395-8 (1999)).

In order to address the problem of being able to select polymerase variants with low catalytic activity or processivity such as may occur along an evolutionary trajectory to a desired phenotype, a variant of CSR, in which only a small region (a "patch") of the gene under investigation is randomized and replicated, is employed. The technique is referred to as "short-patch CSR" (spCSR). spCSR allows for less active or processive polymerases to still become enriched during a round of selection by decreasing the selective advantage given to highly active or processive mutants. This method expands on the previously described method of compartmentalized self-replication, but, because the entire gene is not replicated, the short patch method is also useful for example for investigating specific domains independent of the rest of the protein.

There are many ways to introduce localised diversity into a gene, among these are error-prone PCR (using manganese or synthetic bases, as described above for the Taq polymerase library), DNA shuffling (C. A. Brautigam, T. A. Steitz, *Curr Opin Struct Biol* 8, 54-63 (1998); Y. Li, S. Korolev, G. Waksman, *EMBO J* 17, 7514-25 (1998) cassette mutagenesis (E. Bedford, S. Tabor, C. C. Richardson, *Proc Natl Acad Sci U S A* 94, 479-84 (1997)), and degenerate oligonucleotide directed mutagenesis (Y. Li, V. Mitaxov, G. Waksman, *Proc Natl Acad Sci U S A* 96, 9491-6 (1999); M. Suzuki, D. Baskin, L. Hood, L. A. Loeb, *Proc Natl Acad Sci U S A* 93, 9670-5 (1996)) and its variants, e.g. sticky feet mutagenesis (J. L. Jestin, P. Kristensen, G. Winter, *Angew. Chem. Int. Ed.* 38, 1124-1127 (1999)), and random mutagenesis by whole-plasmid amplification (T. Oberholzer, M. Albrizio, P. L. Luisi, *Chem Biol* 2, 677-82 (1995)). Combinatorial alanine scanning (A. T. Haase, E. F. Retzel, K. A. Staskus, *Proc Natl Acad Sci U S A* 87, 4971-5 (1990)) may be used to generate library variants to determine which amino acid residues are functionally important.

Structural (M. J. Embleton, G. Gorochoy, P. T. Jones, G. Winter, *Nucleic Acids Res* 20, 3831-7 (1992)), sequence alignment (D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* 16, 652-656 (1998)), and biochemical data from DNA polymerase I studies reveal regions of the gene involved in nucleotide binding and catalysis. Several possible regions to target include regions 1 through 6, as discussed in (D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* 16, 652-656 (1998))

(regions 3, 4, and 5 are also referred to as Motif A, B, and C, respectively, in *Taq* DNA polymerase I). Other possible targeted regions would be those regions conserved across several diverse species, those implicated by structural data to contact the nucleotide substrate or to be involved in catalysis or in proximity to the active site, or any other region important to polymerase function or substrate binding.

During a round of selection, each library variant is required to replicate only the region of diversity. This can be easily achieved by providing primers in a PCR reaction which flank the region diversified. CSR selections would be done essentially as described. After CSR selection the short region which is diversified and replicated now is reintroduced into the starting gene (or another genetic framework e.g. a library of mutants of the parent gene, a related gene etc.) using either appropriately situated restriction sites or PCR recombination methods like PCR shuffling or Quickchange mutagenesis etc. The spCSR cycle may be repeated many times and multiple regions could be targeted simultaneously or iteratively with flanking primers either amplifying individual regions separately or inclusively.

To increase stringency in selections at a later stage spCSR is tunable simply by increasing the length of replicated sequence as defined by the flanking primers up to full length CSR. Indeed, for selection for processivity i.e. it may be beneficial to extend the replicated segment beyond the encoding gene to the whole vector using strategies analogous to iPCR (inverted PCR).

spCSR can have advantages over full length CSR not only when looking for polymerase variants with low activities or processivities but also when mapping discrete regions of a protein for mutability, e.g. in conjunction with combinatorial alanine scanning (A. T. Haase, E. F. Retzel, K. A. Staskus, *Proc Natl Acad Sci U S A* 87, 4971-5 (1990)) to determine which amino acid residues are functionally important. Such information may be useful at a later stage to guide semi-rational approaches, i.e. to target diversity to residues /regions not involved in core polymerase activity. Furthermore spCSR may be used to transplant polypeptide segments between polymerases (as with immunoglobulin CDR grafting). A simple swap of segments may lead initially to poorly active polymerases because of steric clashes and may require "reshaping" to integrate segments functionally. Reshaping

may be done using either full length CSR (e.g. from existing random mutant libraries) or spCSR targeted to secondary regions ("Vernier zone" in antibodies).

Short patches may also be located at either N-or C-terminus as extensions to existing polymerase gene sequences or as internal insertions. Precedents for such phenotype modifying extensions and insertions exist in nature. For example both a C-terminal extension of T5 DNA pol and the thioredoxin-binding insertion in T7 DNA pol are critical for processivity in these enzymes and enable them to efficiently replicate the large (> 30kb) T-phage genomes. N-or C-terminal extensions have also been shown to enhance activity in other enzymes.

Example 26: Low temperature CSR using Klenow fragment

Klenow fragment was cloned from E.coli genomic DNA into expression vector pASK75 (as with Taq) and expressed in E.coli strain DH5 $\alpha$ Z1 (Lutz R. & Bujard H. (1997), Nucleic Acids Res 25, 1203). Cells were washed and resuspended in 10mM Tris pH7.5.  $2 \times 10^8$  resuspended cells (20 $\mu$ l) were added to 200 $\mu$ l low temperature PCR buffer (LTP) (Iakobashvili, R. & Lapidot, A. (1999), Nucleic Acids Res., 27, 1566) and emulsified as described (Ghadessy et al. (2001), PNAS, 98, 4552). LTP was 10mM Tris (pH7.5), 5.5M L-proline, 15% w/v glycerol, 15mM MgCl<sub>2</sub> + suitable primers (because proline lowers melting temperature, primers need to be 40-mers or longer) and dNTP's and emulsified as described. Low temperature PCR cycling was 70°C 10min, 50x (70 °C 30sec, 37 °C 12min). Aqueous phase was extracted as described and purified selection products reamplified as described (Ghadessy et al., (2001) PNAS, 98, 4552).

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are apparent to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Primer	Designation	Sequence (5' to 3')
Primer 1	TaqbaXba	GGCGACTCTAGATAACGAGGGCAAAAAATG CGTGGTATGCTTCCTCTTTTGAGCCCAAGGG
Primer 2	TaqfoSal	GCGGTGCGGAGTCGACTCACTCCTTGGCGGA GAGCCAGTCCTC
Primer 3	88ba4	AAAAATCTAGATAACGAGGGCAA
Primer 4	88fo2	ACCACCGAACTGCGGGTGACGCCAAGCG
Primer 5	Taqba(scr)	GGGTACGTGGAGACCCTCTTCGGCC
Primer 6	LMB2	GTAAAACGACGGCCAGT
Primer 7	LMB3	CAGGAAACAGCTATGAC
Primer 8	88ba4LMB3	CAGGAAACAGCTATGACAAAAATCTAGATAA CGAGGGCAA
Primer 9	88fo2LMB2	GTAAAACGACGGCCAGTACCACCGAACTGCG GGTGACGCCAAGCG
Primer 10	LMB388ba5 WA	CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GA (A-G mismatch)
Primer 11	8fo2WC	GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCC (C-C mismatch)

Primer 12	LBA23	GGAGTAGATGCTTGCTT TTCTGAGCC
Primer 13	LFO46	GCTCTGGT TATCTGCATC ATCGTCTGCC

Table 3. Primer sequences used in Examples

## SEQUENCES

Thermostable clone T7-88: Nucleotide sequence

AACCTTGGTATGCTTCCTCTTTTGGAGCCCAAGGGTCGCGTCCTCCTGGTGGACGGCCACCACCTGG  
5 CCTACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGAGCCGGTGCAGGCGGTCT  
ACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTG  
ACGCCAAGGCCCCCTCCTCCCGCCACGAGGCCTACGGGGGTACAAGGCGGGCCGGGCCCCACGC  
CGGAGGACTTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCTGGCGCGCC  
TCGAGGTCCCGGGCTACGAGGCGGACGACGTCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAG  
10 GGCTACGAGGTCCGCATCCTCACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCATCCACG  
TCCTCCACCCCGAGGGGTACCTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCCG  
ACCA GTGGGCGGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCA  
TCGGGGAGAAGACGGCGAAGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCGAGAAC  
CTGGACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCACACGGACGATCTGAAGCTCTCC  
15 TGGGACCTGGCCAAGGTGCGCACCGACCTGCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCC  
GACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTGGCAGCCTCCTCCACGAGTTCGGC  
CTTCTGGAAGCCCCAAGGCCCTGGAGGAGGCCCTGGCCCCCGCCGAAGGGGCCCTTCGTGGGC  
TTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCGGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGC  
CGGGTCCACCGGGCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGGGGGCTTCTC  
20 GCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATG  
CTCCTCGCTACCTCCTGGACCCTTCCAACACACCCCGAGGGGGTGGCCCGGCGCTACGGCGGG  
GAGTGGACGGAGGAGGCGGGGAGCGGGCCGCCCTTCCGAGAGGCTCTTCGCCAACCTGTGGGG  
GAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCTTTCCGCTGT  
CCTGGCCACATGGAGGCCACGGGGGTGCGCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGA  
25 GGTGGCCGAGGAGATCGCCCGCTCGAGGCGGAGGTCTTCGGCTGGCCGGCCACCCCTTCAACCT  
CAACTCCCGGGACGAGCTGGAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGAC  
GGAGAAGACCGCAAGCGCTCCACCAGCGCGCGTCTCTGGAGGCCCTCCGCGAGGCCACCCCAT  
CGTGGAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCC  
GGACCTCATCACCCAGGACGGGCGGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG  
30 CAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGACCCCGCTTGGGCAGAGGAT  
CCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGTCTGGAATATAGCCAGATAGAGCT  
CAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACAT  
CCACACGGAACCGCCAGCTGGATGTTCCGGCGTCCCCCGGGAGGCCGTGGACCCCTGATGCGCCG  
GGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGCCACCGCCTCTCCAGGAGCT  
35 AGCCATCCCTTACGAGGAGGCCAGGCCCTTCAATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCG  
GGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCC



GTCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATG  
GCCTTCAACATGCCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTC  
CCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTCAGGTCCACGACGAGCTGGTCCTCGAGGCC  
CCAAAAGAGAGGGGCGGAGGCCGTGGCCCGGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCCCCT  
5 GGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCTGCCAAGGAGTGAG

Thermostable clone T7-88: Amino Acid Sequence

10 MLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVPQAVYGFAKSLLKALKEDGDAVIVVFDKAP  
SSRHEAYGGYKAGRAPTPEDFRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRLT  
ADKDLYQLLSDRHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTAKKLL  
EEWGSLEALLENLDRPKPAIREKILAHTDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEF  
15 GSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLK  
EARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFA  
NLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPF  
NLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI  
HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIABEGWLLVVDYSQIELRVLAHLSG  
20 DENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPEYEEAQAFI  
ERYFQSFPKVRAWIEKTLLEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTADL  
MKLAMVKLFPRLEEMGARMMLQVHDELVLEAPKERAЕAVARLAKEVMEGVYPLAVPLEVEVGIGED  
WLSAKE\*

25

Thermostable clone T9: Nucleic Acid Sequence

GATGCTCCCTCTTTTGAGCCCAAGGGTCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGC  
ACCTTCCACGCCCTGAAGGGCCTCACCACAGCCGGGGGAGCCGGTGACGGCGGTCTACGGCTTC  
GCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGACGCCAAG  
30 GCCCCCTCCTTCCGCCACGAGGCCTACGGGGGGTACAAGGCGGGCCGGGCCCCACGCCGAGGAC  
TTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCTGGCGCGCCTCGAGGTC  
CCGGGCTACGAGGCGGACGACGTCCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGA  
GGTCCGCATCCTCACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCATCCAGTCCTCCAC  
CCCGAGGGGTACCTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCGACCACTGG  
35 GCCGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGGGGA  
GAAGACGGCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACC

GGCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCTGGGACC  
TGGCCAAGGTGCGCACCGACCTGCCCCTGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGG  
GAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGCTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTG  
GAAAGCCCCAAGGCCCTGGAGGAGGCTCCTGGCCCCCGCGGAAGGGGCCTTCGTGGGCTTTGTG  
5 CTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCCGGCCAGGGGGGGCGGGGTC  
CACCGGGCCCCGAGCCTTATAAAGCCCTCAGAGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA  
GACCTGAGCGTTCTGGCCCTGAGGGAAGGCCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTC  
GCCTACCTCCTGGACCCTTCCAACACCACCCCGAGGGGTGGCCCGGCGCTACGGCGGGGAGTGG  
ACGAGGAGGCGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCT  
10 TGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCCTTCCGCTGTCTGGC  
CCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGC  
CGAGGAGATCGCCCGCCTCGAGGCGAGGTCTTCCGCTGGCCGGCCACCCCTTCAACCTCAACTC  
CCGAGACCAGCTGGAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAA  
GACCGGCAAGCGCTCCACCAGCGCCCGCTCCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGA  
15 GAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCT  
CATCCACCCAGGACGGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCT  
AAGTAGCTCCGATCCCAACCTCCAGAACATCCCGTCCGCAACCCGCTTGGGCAGAGGATCCGCCG  
GGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGT  
GCTGGCCACCTCTCCGGGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACAC  
20 GGAGACCGCCAGCTGGATGTTCCGGCTCCCCGGGAGGCCGTGGACCCCTGATGCGCCGGGCGGC  
CAAGACCATCAACTTCGGGGTCTCTACGGCATGTGGCCACCGCCTCTCCAGGAGCTAGCCATCC  
CTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGA  
TTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCGCGCGCCG  
TACGTGCCAGACCTAGAGGCCCGGTGAAGAGCGTGGGGAGGCGGCCGAGCGCATGGCCTTCAA  
25 CATGCCCGTCCAGGGCACCGCCGCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCT  
GGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCTCGAGGCCCCAAAAG  
AGAGGGCGGAGGCCGTGGCCCGGTGGCCAAGGAGGTATGGAGGGGTGTATCCCTGGCCGTG  
CCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGGGAGTGCACCTGCA  
GGCAGCGCTTGGCGTCACCCGAGTTCGGTGGTACTGGCCGTGTTTTACANN  
30

Thermostable clone T9: Amino Acid Sequence

35 MLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVPQAVYGFAKSLKALKEDGDAVIVVFDKAP  
SFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILT  
ADKDLYQLLSDRHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGKKTARKLL

EEWGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLE  
 LGSLLHEFGLLESPKALEEASWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDL  
 KEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLF  
 ANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHP  
 5 FNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLI  
 HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRRRAFAIEEGWLLVALDYSQIELRVLAHLSG  
 DENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFI  
 ERYFQSFQKVRWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSUREAAERMAFNMPVQGTADL  
 MKLAMVKLFPRLEEMGARMMLQVHDELVLEAPKERAEEAVARLAKEVMEGVYPLAVPLEVEVGIGED  
 10 WLSAKE

# Thermostable clone T13: Amino Acid Sequence

15 MLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVPQAVYGFASLLKALKEDGDAVIVVFDKAP  
 SFRHEAYGGYKAGRAPTEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILT  
 ADKDLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTAKKLL  
 EEWGSLEALLENLDRKPAIREKILAHTDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEF  
 20 GSSLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLK  
 EARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFA  
 NLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPF  
 NLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLI  
 HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRRRAFAIEEGWLLVLDYSQIELRVLAHLSG  
 25 DENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFI  
 ERYFQSFQKVRWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSUREAAERMAFNMPVQGTADL  
 MKLAMVKLFPRLEEMGARMMLQVHDELVLEAPKERAEEAVARLAKEVMEGVYPLAVPLEVEVGIGED  
 WLSAKE

# 30 Thermostable clone 8 (T8): Nucleic Acid Sequence

TCGTGGTACGCATCCTCTTTTGTAGCCCAAGGGCCGCTCCTCCTGGTGACGGCCACCACTGGCC  
 TACCGCACCTTCCACGCCCTGAAGGGCCTCACCACAGCCGGGGGAGCCGGTGACGGCGGTCTAC  
 GGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGAC  
 35 GCCAAGGCCCTCCTCCCGCCACGAGGCCTACGGGGGTACAAGGCGGGCCGGGCCCCACGCCG

GAGGACTTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCTGGCGCGCCTC  
GAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGG  
CTATGAGGTCCGCATCCTCACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCATCCACGTC  
CTCCACCCCGAGGGGTACCTCATCACCCCGGCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGAC  
5 CAGTGGGCGGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATC  
GGGGAGAAGACGGCGAAGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCGAGAACCT  
GGACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCACACGGACGATCTGAAGCTCTCCTG  
GGACCTGGCCAAGGTGCGCACCGACCTGCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCG  
ACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTTCGGCC  
10 TTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCT  
TTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCCGCCGAGGGGTGGCC  
GGGTCCACCGGGCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCG  
CCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGCGACGACCCCATGC  
TCCTCGCTACCTCCTGGACCCCTTCCAACACCAACCCCGAGGGGGTGGCCCGCGCTACGGCGGGG  
15 AGTGGACGGAGGAGGCGGGGAGCGGGCCGCTTTCCGAGAGGCTTTCGCCAACCTGTGGGGG  
AGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCTTTCCGCTGTC  
CTGGCCACATGGAGGCCACAGGGGTGCGCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAG  
GTGGCCGAGGAGATCGCCCGCCTCGAGGCGGAGGTCTTCCGCTGGCCGCCACCCCTTCAACCTC  
AACTCCCGGGACCAAGCTGGAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG  
20 GAGAAGACCGGAAGCGCTCCACCAGCGCCGCGTCTGGAGGCCCTCCGCGAGGCCACCCCATC  
GTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCG  
GACCTCATCCACCCAGGACGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGC  
AGGCTAAGTAGCTCCGATCCCAACCTCCAGAATCCCGTCCGCACCCCGCTTGGGCAGAGGATC  
CGCCGGGCTTTCATCGCCGAGGAGGGGTGGCTATTGGTGGTCTGGACTATAGCCAGATAGAGCTC  
25 AGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATC  
CACACGAAACCGCCAGCTGGATGTTTCGGCGTCCCCGGGAGGCCGTGGACCCCTGATGCGCCGG  
GCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTGCGCCACCGCCTCTCCAGGAGCTA  
GCCATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGG  
GCCTGGATTGAGAAGACCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCG  
30 CCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGG  
CCTTCAACATGCCCGTCCAGGGCACCGCCCGGACCTCATGAAGCTGGCTATGGTGAAGCTCTCCC  
CAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTCAGGTCCACGACGAGCTGGTCTCGAGGCCCC  
AAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTATGGAGGGGGTGTATCCCCTGG  
CCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGAGT  
35

Thermostable clone 8 (T8): Amino Acid Sequence

PLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVPQAVYGFASLLKALKEDGDAVIVVFDKAPSSR  
HEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRLTAD  
KDLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPVKGIGEKTAKKLEEL  
WGSLEALLENLDRPKPAIREKILAHTDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGS  
5 LLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEA  
RGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANL  
WGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNL  
NSRDQLERVLFDELGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHP  
TGR LHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIAEEGWLLVVL DYSQIELRVLAHLSGDE  
10 NLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPEYEEAQAFIER  
YFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKS VREAAERMAFNMPVQGTADLMK  
LAMVKLFPRL EEMGARMLLQVHDEL VLEAPKERA EAVARLAKEVM EGVYPLAVPLEVEVGIGEDWLS  
AKE\*

15 Note: First two amino acids at N terminus not sequenced.

#### Heparin Resistant Clone 94: Nucleic Acid Sequence

20 ATTTTGTAGCCCAAGGGCCGCTCCTCCTGGTGACGGCCACCACCTGGCCTACCGCACCTTCCACG  
CCCTGAAGGGCCTCACCACCAGCCGGGGGAGCCGGTGACGGCGGTCTACGGCTTCGCCAAGAGCC  
TCCTCAAGGCCCTCAAGGAGGACGGGGACGGGTGATCGTGGTCTTTGACGCCAAGGCCCTCCT  
TCCGCCACGAGGCCTACGGGGGTACAAGGCGGGCCGGGCCCCACGCCGAGGACTTTCCCGGGC  
25 AACTCGCCCTCATCAAGGAGCTGGTGACCTCCTGGGGCTGGCGCGCCTCGAGGTCCCGGGCTACG  
AGGCGGACGACGTCCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATC  
CTACCGCCGACAAAGACCTTTACCAGTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGT  
ACCTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAAGTGGGCCGACTACC  
GGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGTGTCAAGGGCATCGGGGAGAAGACGGCG  
30 AGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCGGCTGGAGCC  
CGCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGT  
GCGCACCGACCTGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTA  
GGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCA  
AGGCCCCGGAGGAGGCCCTGGCCCCCGCGGAAGGGGCCCTTCGTGGGCTTTGTGCTTTCCCGCA  
35 AGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCC  
CCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCG  
TTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGACGACCCCATGCTCCTCGCCTACCTCCT

GGACCCCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGG  
CGGGGGAGCGGGCCGCCCTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAG  
GAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCTTCCGCTGTCCTGGCCACATGGAG  
GCCACGGGGGTGCGCCTGGACGTGTCCTATCTCAGGGCCTTGTCGCGGGAGGTGGCCGAGGAGATC  
5 GCGCGCCTCGAGGCCGAGGTCTTCCGCTGGCCGCCACCCCTTCAACCTCAACTCCCGGGACCAG  
CTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCGCCATCGGCAAGACGGAGAAGACCGGCAAG  
CGCTCCACCAGCGCCGCGTCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCTG  
CAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCC  
AGGACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCC  
10 GGTCCCAACCTCCAGAGCATCCCCGTCCGACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATC  
GCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCAC  
CTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCC  
AGCTGGATGTTCCGGCTCCCCCGGGAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATC  
AACTTCGGGGTCTCTACGGCATGTCCGCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGG  
15 AGGCCAGGCCTTCATTGAGCGCTACTTTAGAGCTTCCCAAGGTGCGGGCCTGGATTGAGAAGA  
CCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCCGCGCTACGTGCCA  
GACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGT  
CCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCAGGCTGGAGGAAAT  
GGGGGCCAGGATGCTCTTCAGGTCCACGACGAGCTGGTCTCGAGGCCCAAAAGAGAGGGCGG  
20 AGGCCGTGGCCCGCTGGCCAAGGAGGTCATGGAGGGGTGTATCCCTGGCCGTGCCCTGGAGG  
TGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGATT

25 Heparin Resistant Clone H94. Amino Acid Sequence

FEFKGRVLLVDGHHLAYRTFHALKGLTTSRGEFVQAVYGFAKSLKALKEDGDAVTVVFDKAPSFRH  
EAYGGYKAGRPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADK  
DLYQLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGECTARKLLEEW  
30 GSLEALLKNLDRLEPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSL  
LHEFGLLSPKAPPEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEAR  
GLLAKDLSVLALREGLPDPDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERLFANLW  
GRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVSYLRLSREVAEEIARLEAEVFRLAGHPFNLS  
RDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLPDIHPRT  
35 GRLHTRFNQTATATGRLSSSGPNLQSIPVRTPLGQRIRRAFIABEGWLLVALDYSQIELRVLAHLSGDNEL  
IRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPIYEEAQAFIERYF  
QSFQKVRWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVRERAAERMAFNMPVQGTAAADLMKLA

MVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAK  
E\*

Note: N-TERMINAL 5 amino acids not determined.

5

### Heparin Resistant Clone 15: Nucleic Acid Sequence

10 TTTGAGCCCAAGGGCCGCTCCTCCTGGTGGACGGCCACCACTGGCCTACCGCACCTTCCACGCCC  
TGAAGGGCCTACCAACCAGCCGGGGGAGCCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCC  
TCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGACGCCAAGGCCCTCCTTCC  
GCCACGAGGCCTACGGGGGTACAAGGCGGGCCGGGCCCCACGCCGGAGGACTTTCCCCGGCAA  
CTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCTGGCGCGCTCGAGGTCCCGGGCTACGAG  
15 GCGGACGACGTCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCT  
CACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCTCCACCCCGAGGGGTAC  
CTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAAGTGGGCCGACTACCGG  
GCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGTGTCAAGGGCATCGGGGAGAAGACGGCGAG  
GAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCGGCTGGAGCCCG  
20 CCATCCGGGAGAAGATCCTGGCCACATGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGC  
GCACCGACCTGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGG  
GCCTTCTGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAGCCCCAAG  
GCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAG  
GAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGTCCGGGTCCACCGGGCCCCC  
25 GAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTT  
CTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCTACCTCCTGG  
ACCTTTCCAACACCAACCCCGTGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCG  
GGGGAGCGGGCCGCCCTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGA  
GAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCTTTCCGCTGTCTGGCCACATGGAGGC  
30 TACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGC  
CCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTG  
GAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGC  
TCCACCAGCGCCGCGTCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCTGCAG  
TACCGGGAGCTCACCAGGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCACCCCAAG  
35 ACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGGT  
CCCAACCTCCAGAGCATCCCCGTCCGACCCCGCTTGGGCAGAGGATCGCCGGGGCCTTCATCGCC  
GAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTC

TCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAG  
CTGGATGTTCCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAA  
CTTCGGGGTCTCTACGGCATGTCGGCCCCACCGCCTCTCCAGGAGCTAGCCATCCCTTACGAGGAG  
GCCCAGGCCTTCATTGAGCGCTACTTTAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAGAAGACC  
5 CTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCCGCTACGTGCCAGA  
CCTAGAGGCCCGGGTGAAGAGCGTGGGGAGGCGCCGAGCGCAGGGCCTTCAACATGCCCGTCC  
AGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGG  
GGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCTCGAGGCCCCAAAAGAGAGGGCGGAG  
GCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTG  
10 GAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGAGT

15 Heparin Resistant Clone 15: Amino Acid Sequence

PLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVQAVYGFASLLKALKEDGDAVIVVFDKAPSFR  
HEAYGGYKAGRPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTAD  
KDLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGETARKLLEE  
20 WGSLEALLKNLDRLEPAIREKILAHMDDLKLSWDLAKVRTDPLEVDFAKRREPDRERLRAFLEFEG  
SLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGRVHRAPEPYKALRDLKE  
ARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPVGVARRYGGEWTEEAGERAAALSERLFAN  
LWGRLEGEERLLWLRYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFN  
LNSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTRLKSTYIDPLDLIHP  
25 RTGRLHTRFNQTATATGRLSSGPNLQSIPTVTPLGQIRRAFIIEGWLLVALDYSQIELRVLAHLSGDE  
NLIRVFQEGRIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHLRSQELAIPYEEAQAFIER  
YFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSUREAAERRAFNMPVQGTAAADLMKL  
AMVKLFPRLEEMGARMMLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSA  
KE\*

30

Note: N-terminal 5 amino acids not determined.

Mismatch extension clone M1: Nucleic acid sequence.

TTGGAATGCTCCCTCTTTTGGAGCCCAAAGGCCGCGTCTCTCTGGTGGACGGCCACCACCTGGCCTA  
35 CCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGAGCCGGTGCAGGCGGTCTACGG



CTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGACGC  
CAAGGCCCCCTCCTTCCGCCACGAGGCCTACGGGGGTACAAGGCGGCCCGGGCCCCACGCCGGA  
GGACTTTCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGATCTCCTGGGGCTGGCGCGCCTCGA  
GGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGGCT  
5 ACGAGGTCCGCATCCTCACCGCCGACAAAGGCCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCT  
CCACCCCGAGGGGTACCTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCA  
GTGGGCCGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGG  
GGAGAAGACGGCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGG  
ACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCCTGGG  
10 ATCTGGCCAAGGTGCGCACCGACCTGCCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACC  
GGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTC  
TGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGAAGGGGCCTTCGTGGGCTTTG  
TCCTTTCCCGCAGGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCAGGGGGGGCCGGG  
TCCACCGGGCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGGGGGGCTTCTCGCCA  
15 AAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCCTCCCGCCCGGCGACGACCCCATGCTCC  
TCGCTACCTCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGTACGGCGGGGAGT  
GGACGGAGGAGGCGGGGAGCGGGCCGCCCTTTCGAGAGGCTCTTCGCCAACCTGTGGGGGAGG  
CTTGAGGGGGAGGAGAGGCTCCTTTGGCTTACCGGGAGGTGGAGAGGCCCTTTCCGCTGTCTG  
GCCCATGAGGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTG  
20 GCCGAGGAGATCGCCCGCTCGAGGCCGAGGTCTTCCGCTGGCCGGCCACCCCTTCAACCTCAAC  
TCCCGGGACAGCTGGAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAG  
AAGACCGGCAAGCGCTCCACCAGCGCCCGCTCCTGGGGGGCCCTCCGCGAGGCCACCCCATCGTG  
GAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTACCGGAC  
CTCATCCACCCAGGACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGG  
25 CTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCGTCCGACCCCGCTTGGGCAGAGGATCCGC  
CGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGTCTGGAATATAGCCAGATAGAGCTCAGG  
GTGCTGGCCACCTCTCCGGCAGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCAC  
ACGGAGACCGCCAGCTGGATGTTGGCGTCCCCCGGAGGCCGTGGACCCCTGATGCGCCGGGGC  
GCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTCCGCCACCGCCTCTCCAGGAGCTAGCC  
30 ATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTAGAGCTTCCCAAGGTGCGGGCC  
TGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCCG  
CCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGGGGCGGCCGAGCGCATGGCCT  
TCAACATGCCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCA  
GGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCTCGAGGCCCAA  
35 AAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTGATGGAGGGGGTGTATCCCTGGCC  
GTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGAGTCGACCT  
GCAGGCAGCGCTTGGCGTCACCCGAGTTCGGTGGTTAATAAGCTTGACCTGTGAAGTGAATAATG  
GCGCACATTGTGCGACATTTTTTTGTCTGCCGTTTACCGCTACTGCGTCACGGATCTCCACGCGCC  
CTGTAGCGGCGCATTAAGCGCGGGGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG

CGCCCTAGCGCCCGCTCCTTTTCGCTTCTTCCCTTCCTTTCTCGCCACGTTCCGCCGGCTTTCCCCGTC  
AAGCTCTAAATCGGGGGCTCCCTTTAGGGGTTCCCGATTAGTGCTTTTACGGGACCTCGAACCCAA  
AAAATTGATTAGG

- 5 Mismatch extension clone M1: Amino acid sequence.
- GMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVPQAVYGFAKSLKALKEDGDAVIVVFDAKA  
PSFRHEAYGGYKAARAPTPEDFRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRIL  
TADKGLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGECTARKL  
LEEWGSLEALLKNLDRKPAIRKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERL  
10 EFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRREPMWADLLALAAARGGRVHRAPEPYKALRDL  
KEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLF  
ANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHP  
FNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTAAVLGALREAHPIVEKILQYRELTKLKSTYIDPLPDL  
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIAEEGWLLVVDYSQIELRVLAHLS  
15 GDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRRAKTINFGVLYGMSAHRLSQELAIPEYEAQAF  
IERYFQSPFKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVRGAAERMAFNMPVQGTAADL  
MKLAMVKLFPRL EEMGARMMLQVHDEL VLEAPKERABAVARLAKEVMEGVYPLAVPLEVEVGIGED  
WLSAKE
- 20 Note: N<sup>o</sup>terminal 2 amino acids not determined.

Mismatch extension clone M4: Nucleic acid sequence.

- 25 TCTTTATGAGCCCAAGGGCCGCTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCCAC  
GCCCTGAAGGGCCTCACCACCAGCCGGGGGAGCCGGTGCAGGCGGTCTACGGCTTCGCCAAGAG  
CCTCCTCAAGGCCCTCAAGGAGGGCGGGGACGCGGTGATCGTGGTCTTTGACGCCAAGGCCCCCTC  
CTTCCCCATGAGGCCTACGGGGGGTACAAGGCGGGCCGGGCCCCACGCCGAGGACTTTCCCCG  
ACAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCTGACGCGCCTCGAGGTCCCGGGCTA  
CGAGGGCGGACGACGTCTGGCCAGCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCA  
30 TCCTCACCGCCGACAAAGACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCTCCACCCCGAGGG  
GTACCTCATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCGACCACTGGGCCGACTA  
CCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGGGGAGAAGACGG  
CGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCGGCTGAAG  
CCC GCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCCTGGGACCGGGCCAAG  
35 GTGCGCACCGACCTGCCCCGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCT  
TAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCC

AAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGC  
AAGGAGCCCATGTGGGCCGATCTTCTAGCCCTGGCCGCCAGGGGGGGCCGGGTCCACCGGGCC  
CCCGAGCCTTATAAAGCCCTCGGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGC  
GTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCGACGACGACCCCATGCTCCTCGCCTACCTCC  
5 TGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGAGGAG  
GCAGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGA  
GGAAAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCTTTCCGCTGTCCTGGCCACATGGA  
GGCCACGGGGGTGCGCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGAT  
CGCCCGCCTCGAGGCCGAGGTCTTCGCTGGCCGCCACCCCTTCAACCTCAACTCCCGGGACCA  
10 GCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAA  
GCGCTCCACGAGCGCCGCGTCTGGGGGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCT  
GCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCC  
CAGGACGGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTC  
CGATCCCAACCTCCAGAGCATCCCCGTCCGACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCAT  
15 CGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCA  
CCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCAGGAGGGGCGGGACATCCACACGGAGACCGC  
CAGCTGGATGTTCCGGCTCCCCCGGGAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCAT  
CAACTTCGGGGTCTCTACGGCATGTCCGCCACCGCTCTCCAGGAGCTAGCCATCCCTTACGAG  
GAGGCCCAGGCCTTCATTAAGCGCTACTTTCAGAGCTTCCCAAGGTGCGGGCCTGGATTGAGAAG  
20 ACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCTCTTCGGCCGCCGCCGCTACGTGCC  
AGACCTAGAGGCCCGGTGAAGAGCGTGCGGGAGCCGGCCGAGCGCATGGCCTTCAACATGCCCC  
TCCAGGTACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAA  
TGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCTCGAGGCCCCAAAAGAGAGGGCG  
GAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGTGTATCCCCTGGCCGTGCCCTGGAG  
25 GTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGAGT

Mismatch extension clone M4: Amino acid sequence.

LYEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEVQAVYGFAKSLKALKEGGDAVIVVFDKAPSFP  
HEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLTRLEVPGYEADDVLASLAKKAEKEGYEVRILTADK  
30 DLYQLLSDRHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTAARKLLEW  
GSLEALLKNLDRKPAIRKILAHMDDLKLSWDRAKVRITDLPLEVDFAKRREPDRERLRAFLERLEFGS  
LLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALGDLKEA  
RGLLAKDLSVLALREGLGLPDDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERLFANL  
WGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNL  
35 NSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLGALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHP  
RTGRLHTRFNQTATATGRLSSDPNLQSIPVRTPLGQIRRAFIAGEGWLLVALDYSQIELRVLAHLSDGE  
NLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTNFGVLYGMSAHRLSQELAIPIYEEAQAFIKR

YFQSFPKVRWIEKTLLEGRRRGYVETLFGRRRYVPDLEARVKSVREPAERMAFNMPVQGTAAADLMK  
LAMVKLFPRLEEMGARMLLQVHDELVEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLS  
AKE

- 5 Note: N-terminal 6 amino acids not determined.

CLAIMS

1. A method of selecting a nucleic acid processing (NAP) enzyme, the method comprising the steps of:
  - 5 (a) providing a pool of nucleic acids comprising members encoding a NAP enzyme or a variant of the NAP enzyme;
  - (b) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool together with the NAP enzyme or variant encoded by the nucleic acid member;
  - (c) allowing nucleic acid processing to occur; and
  - 10 (d) detecting processing of the nucleic acid member by the NAP enzyme.
2. A method of selecting an agent capable of modifying the activity of a NAP enzyme, the method comprising the steps of:
  - (a) providing a NAP enzyme;
  - 15 (b) providing a pool of nucleic acids comprising members encoding one or more candidate agents;
  - (c) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool, the agent encoded by the nucleic acid member, and the NAP enzyme; and
  - (d) detecting processing of the nucleic acid member by the NAP enzyme.
- 20 3. A method according to Claim 2, in which the agent is a promoter of NAP enzyme activity.

4. A method according to Claim 2 or 3, in which the agent is an enzyme, preferably a kinase or a phosphorylase, which is capable of acting on the NAP enzyme to modify its activity.
5. A method according to Claim 2 or 3, in which the agent is a polypeptide involved in a metabolic pathway, the pathway having as an end product a substrate which is involved in a nucleic acid processing reaction.
6. A method according to Claim 2 or 3, in which the agent is a polypeptide capable of producing a substrate or consuming an inhibitor in a nucleic acid processing reaction.
7. A method according to Claim 2 or 3, in which the agent is a polypeptide capable of modifying a nucleotide primer or nucleoside triphosphate substrate used in a nucleic acid processing reaction such that
  - a) its 3' end becomes extendable; or
  - b) a substrate portion appended to the nucleotide primer or nucleoside triphosphate is modified such as to allow detection or capture of product appendage of the incorporated nucleotide primer or nucleoside triphosphate.
8. A method of selecting a pair of polypeptides capable of stable interaction, the method comprising:
  - (a) providing a first nucleic acid and a second nucleic acid, the first nucleic acid encoding a first fusion protein comprising a first subdomain of a NAP enzyme fused to a first polypeptide, the second nucleic acid encoding a second fusion protein comprising a second subdomain of a NAP enzyme fused to a second polypeptide; in which stable interaction of the first and second NAP subdomains generates processing activity, and in which at least one of the first and second nucleic acids is provided in

the form of a pool of nucleic acids encoding variants of the respective first and/or second polypeptide(s);

(b) subdividing the pool or pools of nucleic acids into compartments, such that each compartment comprises a first nucleic acid and a second nucleic acid together with  
5 respective fusion proteins encoded by the first and second nucleic acids;

(c) allowing the first polypeptide to bind to the second polypeptide, such that binding of the first and second polypeptides leads to stable interaction of the NAP subdomains to generate NAP enzyme activity; and

(d) detecting processing of at least one of the first and second nucleic acids by the  
10 NAP enzyme.

9. A method of selecting a pair of polypeptides capable of stable interaction, the method comprising:

(a) providing a first nucleic acid and a second nucleic acid, the first nucleic acid encoding a first fusion protein comprising a first subdomain of polypeptide capable of  
15 enhancing the activity of a NAP enzyme fused to a first polypeptide, the second nucleic acid encoding a second fusion protein comprising a second subdomain of polypeptide capable of enhancing the activity of a NAP enzyme fused to a second polypeptide; in which stable interaction of the first and second NAP subdomains generates processing activity, and in which at least one of the first and second nucleic  
20 acids is provided in the form of a pool of nucleic acids encoding variants of the respective first and/or second polypeptide(s);

(b) subdividing the pool or pools of nucleic acids into compartments, such that each compartment comprises a first nucleic acid and a second nucleic acid together with  
respective fusion proteins encoded by the first and second nucleic acids;

(c) allowing the first polypeptide to bind to the second polypeptide, such that binding of the first and second polypeptides leads to stable interaction of the subdomains of the  
25 NAP activity enhancing agent to generate NAP enzyme activity; and

(d) detecting processing of at least one of the first and second nucleic acids by the NAP enzyme.

10. A method of selecting a polypeptide capable of stable folding, the method comprising:

5 (a) providing a pool of nucleic acids comprising members encoding one or more candidate polypeptides fused to a NAP enzyme;

(b) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool, the fusion polypeptide encoded by the nucleic acid member, and the NAP enzyme; and allowing for folding of the fusion polypeptide such that a folded fusion polypeptide will not inhibit NAP activity; and

(c) detecting processing of the nucleic acid member by the NAP enzyme.

11. A method of selecting a polypeptide capable of promoting stable folding, the method comprising:

15 (a) providing a poorly folding polypeptide fused to a NAP enzyme and the candidate chaperone;

(b) providing a pool of nucleic acids comprising members encoding one or more candidate chaperones;

20 (c) subdividing the pool of nucleic acids into compartments, such that each compartment comprises a nucleic acid member of the pool, the candidate chaperone encoded by the nucleic acid member, the NAP-enzyme fusion to the poorly folding polypeptide; and allowing for chaperone-aided folding of the fusion polypeptide such that a folded fusion polypeptide will not inhibit NAP activity

(d) detecting processing of the nucleic acid member by the NAP enzyme.

12. A method according to any preceding claim, wherein the NAP enzyme is a replicase enzyme.



13. A method according to any preceding claim, wherein the NAP enzyme is a replicase enzyme and the processing of the nucleic acid member constitutes replication of the whole or a segment(s) of said member.
- 5 14. A method according to any preceding claim, wherein the NAP enzyme is a replicase enzyme and the processing of the nucleic acid member comprises either a fill-in reaction of a 5' overhang appended to said member or an extension of a 3' end of said member.
15. A method according to claim 12, in which amplification of the nucleic acid results from more than one round of nucleic acid replication.
- 10 16. A method according to any one of claims 12 to 15, in which the amplification of the nucleic acid is an exponential amplification.
- 15 17. A method according to any one of claims 12 to 16, in which the amplification reaction is a polymerase chain reaction (PCR), a reverse transcriptase-polymerase chain reaction (RT-PCR), a nested PCR, a ligase chain reaction (LCR), a transcription based amplification system (TAS), a self-sustaining sequence replication (3SR), NASBA, a transcription-mediated amplification reaction (TMA), or a strand-displacement amplification (SDA).
- 20 18. A method according to any of Claims 12 to 17 as dependent upon claim 1, in which the post-amplification copy number of the nucleic acid member is substantially proportional to the activity of the replicase.
19. A method according to any one of Claims 12 to 17 as dependent upon claim 2, in which the post-amplification copy number of the nucleic acid member is substantially proportional to the activity of the agent.

20. A method according to any of Claims 12 to 17 as dependent upon claims 3 to 5, in which the post-amplification copy number of the nucleic acid member is substantially proportional to the binding affinity and/or kinetics of the first and second polypeptides.
- 5 21. A method according to any one of claims 12 to 20, in which nucleic acid replication is detected by assaying the copy number of the nucleic acid member.
22. A method according to any one of claims 12 to 20, in which nucleic acid replication is detected by assaying the copy number of segments of the nucleic acid member.
23. A method according to any one of claims 12 to 20, in which nucleic acid replication is detected by assaying the presence of tagging of the nucleic acid member.
- 10 24. A method according to any one of claims 12 to 23, in which nucleic acid replication is detected by determining the activity of a polypeptide encoded by the nucleic acid member.
- 15 25. A method according to any preceding claim, in which the conditions in the compartment are adjusted to select for a NAP enzyme or an agent active under such conditions, or a pair of polypeptides capable of stable interaction under such conditions.
26. A method according to any one of claims 12 to 24, in which the replicase activity is a templated replicase activity such as a polymerase, reverse transcriptase or ligase activity.
- 20 27. A method according to any preceding claim, in which the polypeptide is provided from the nucleic acid by *in vitro* transcription and translation.
28. A method according to any of Claims 1 to 18, in which the polypeptide is provided from the nucleic acid *in vivo* in an expression host.

29. A method according to any preceding claim, in which the compartments comprise aqueous compartments of a water-in-oil emulsion.
30. A method according to Claim 29, in which the water-in-oil emulsion is produced by emulsifying an aqueous phase with an oil phase and a surfactant comprising 4.5% v/v Span80, 0.4% v/v Tween80 and 0.1% v/v TritonX100, or a surfactant comprising Span80, Tween80 and TritonX100 in substantially the same proportions.
31. A NAP enzyme identified by a method according to any preceding claim.
32. A NAP enzyme according to Claim 31, which has a greater thermostability than a corresponding unselected enzyme.
33. A NAP enzyme according to Claim 31 or 32, which is a *Taq* polymerase having more than 10 times increased half-life at 97.5 °C when compared to wild type *Taq* polymerase.
34. A *Taq* polymerase mutant comprising the mutations: F73S, R205K, K219E, M236T, E434D and A608V.
35. A NAP enzyme according to Claim 31, which inhibited to a lesser extent by heparin than is a corresponding unselected enzyme.
36. A NAP enzyme according to Claim 31 or 35, which is a *Taq* polymerase active at a concentration of 0.083 units/ $\mu$ l or more of heparin.
37. A *Taq* polymerase mutant comprising the mutations: K225E, E388V, K540R, D578G, N583S and M747R.

38. A NAP enzyme according to Claim 31 which is a replicase enzyme and is capable of extending a primer having a 3' mismatch.
39. A NAP enzyme according to Claim 31 which is a replicase enzyme and is capable of extending a primer having a 3' unnatural base (e.g. 5-nitroindole, or 3-carboxyamido-5-nitroindole).
40. A NAP enzyme according to Claim 31 which is a replicase enzyme and has an enhanced capability to utilize  $\alpha$ -thio dNTP's as nucleotide substrates.
41. A NAP enzyme according to Claim 31 which is a replicase enzyme and has an enhanced capability to replicate substrates 25kb in size in the absence of processivity factors or a 3'5'exonuclease proof-reading domain.
42. A replicase enzyme according to any one of claims 38 to 41, in which the 3' mismatch is a 3' purine-purine mismatch or a 3' pyrimidine-pyrimidine mismatch.
43. A replicase enzyme according to any one of claims 38 to 42 in which the 3' mismatch is an A-G mismatch or in which the 3' mismatch is a C-C mismatch.
44. A *Taq* polymerase mutant comprising the mutations: G84A, D144G, K314R, E520G, check F598, A608V, E742G
45. A *Taq* polymerase mutant comprising the mutations: D58G, R74P, A109T, L245R, R343G, G370D, E520G, N583S, E694K, A743P
46. A water-in-oil emulsion obtainable by emulsifying an aqueous phase with an oil phase in the presence of a surfactant comprising 4.5% v/v Span80, 0.4% v/v Tween80 and 0.1% v/v TritonX100, or a surfactant comprising Span80, Tween80 and TritonX100 in substantially the same proportions.

Self-evolving polymerase:  
Linking gene copy number to enzymatic turnover

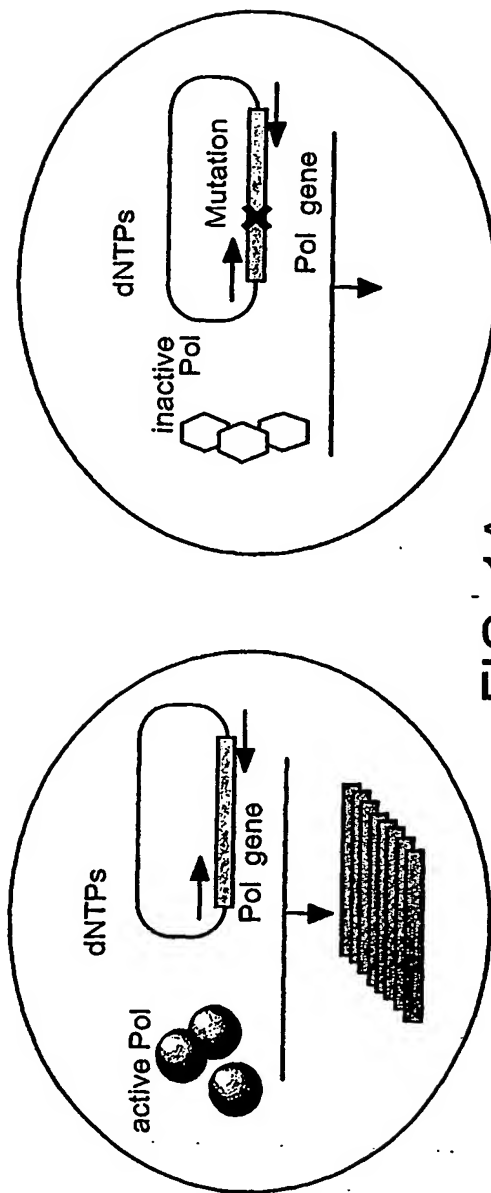
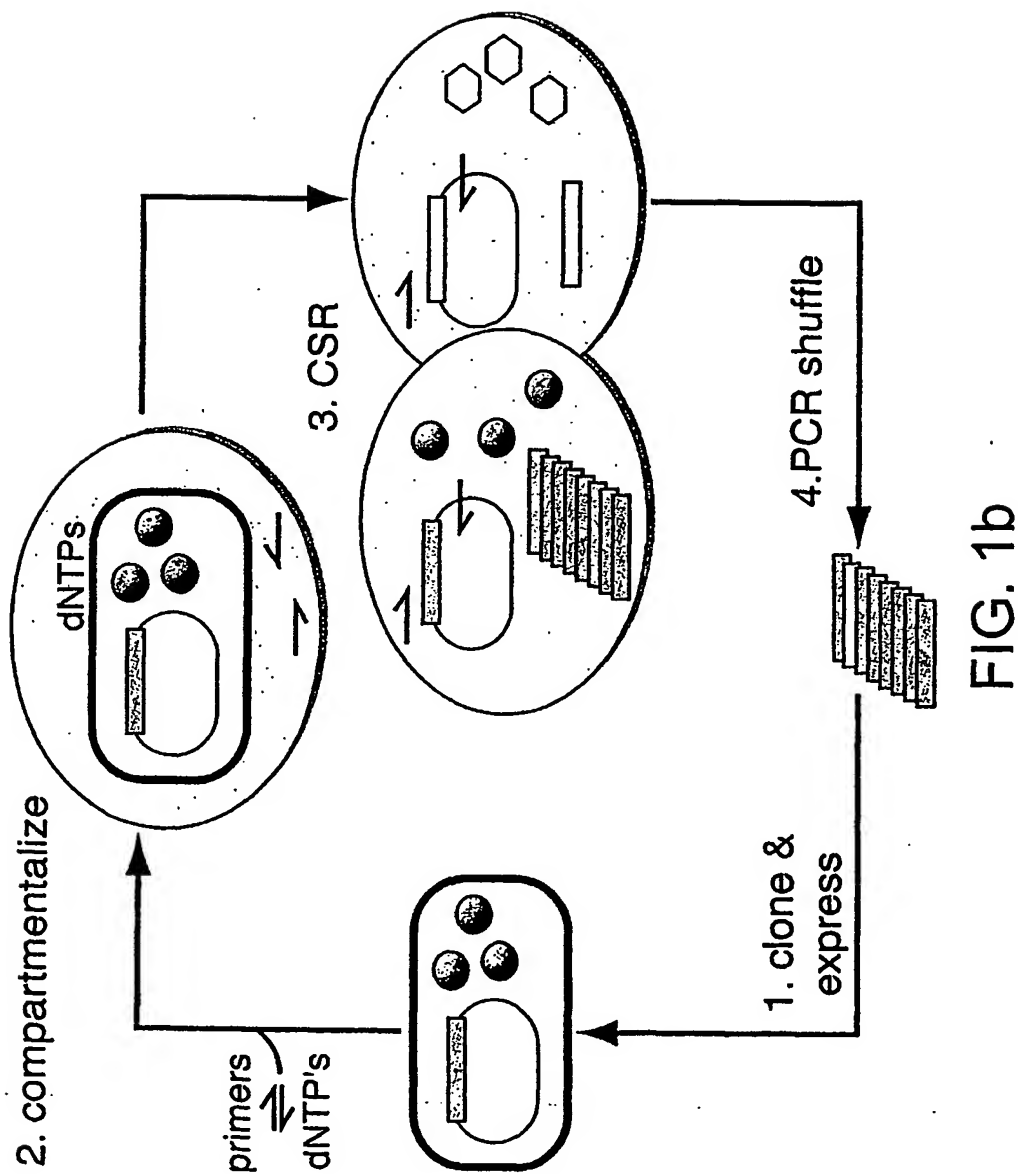


FIG. 1A



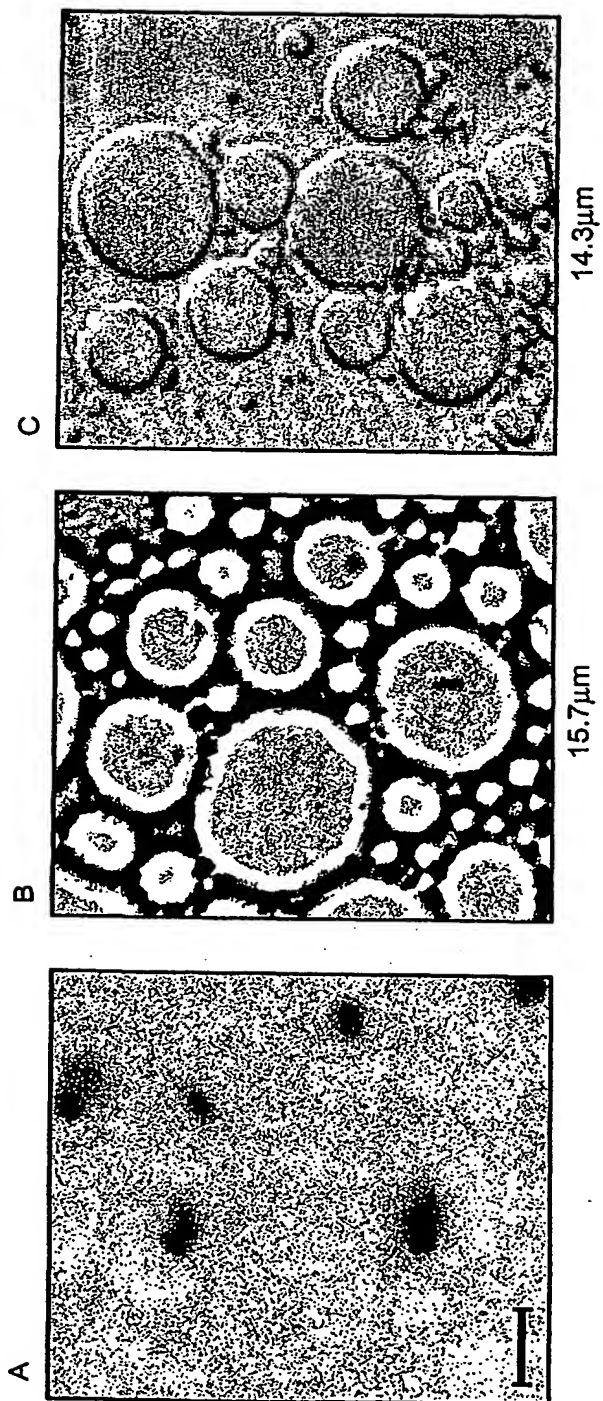


FIG. 2

4/9

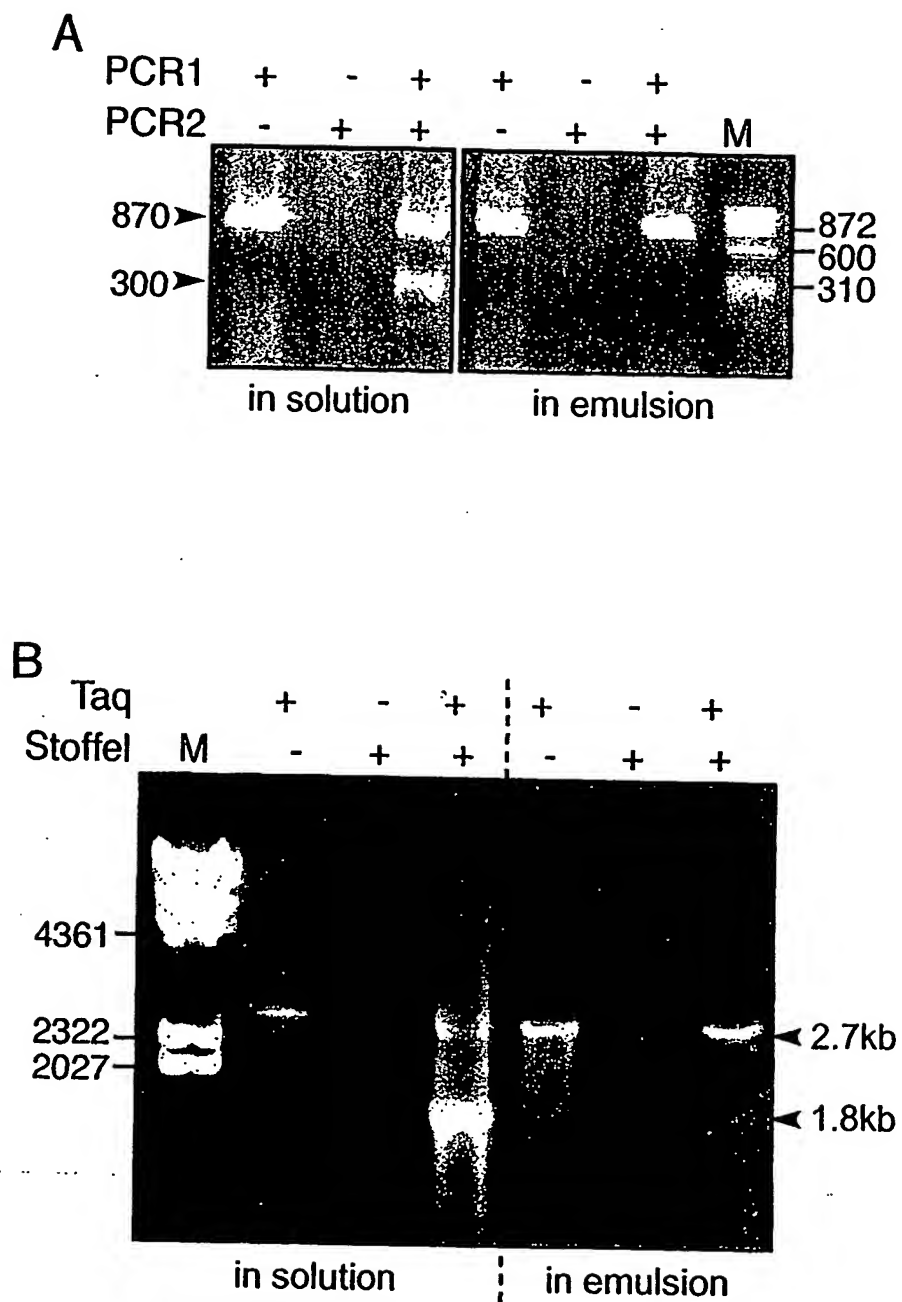


FIG. 3



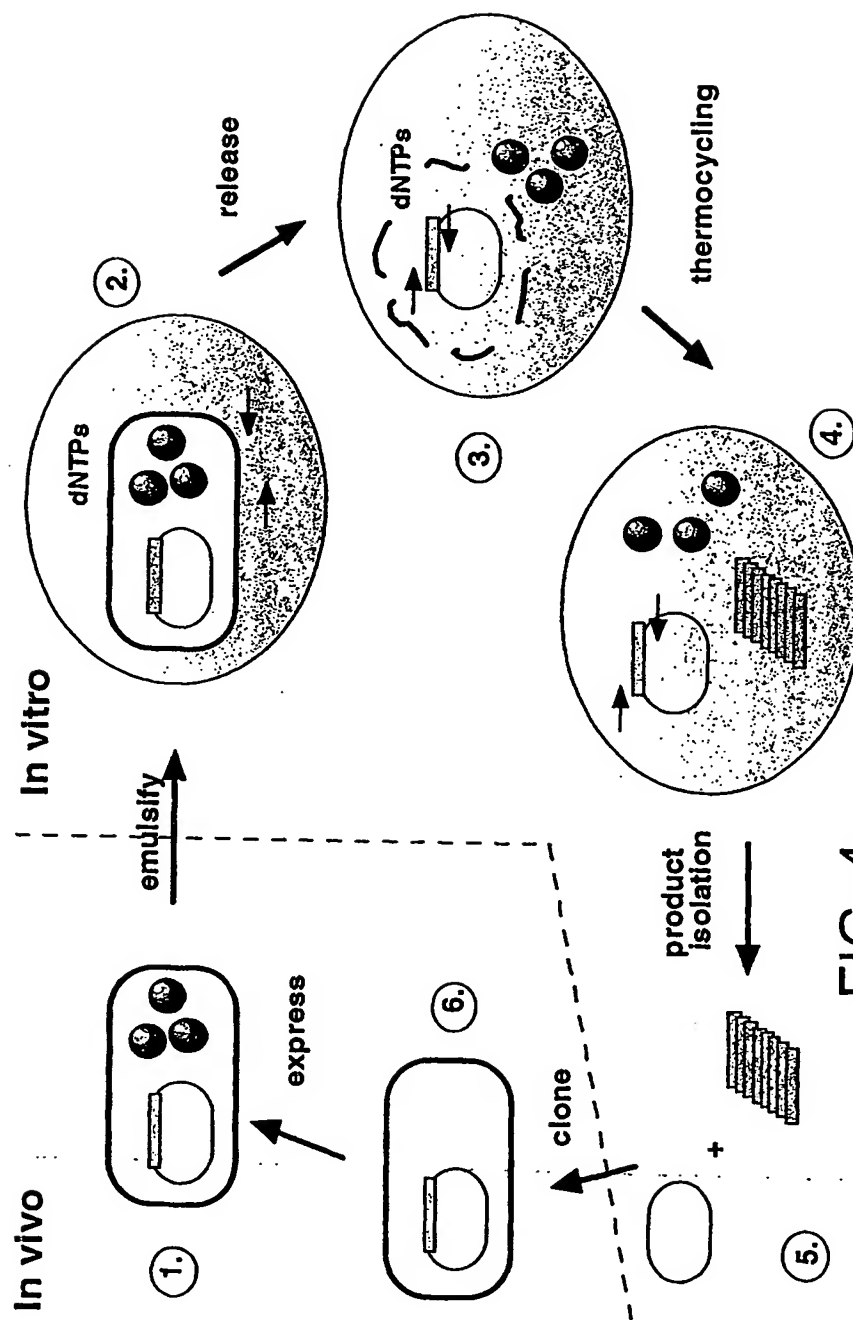


FIG. 4

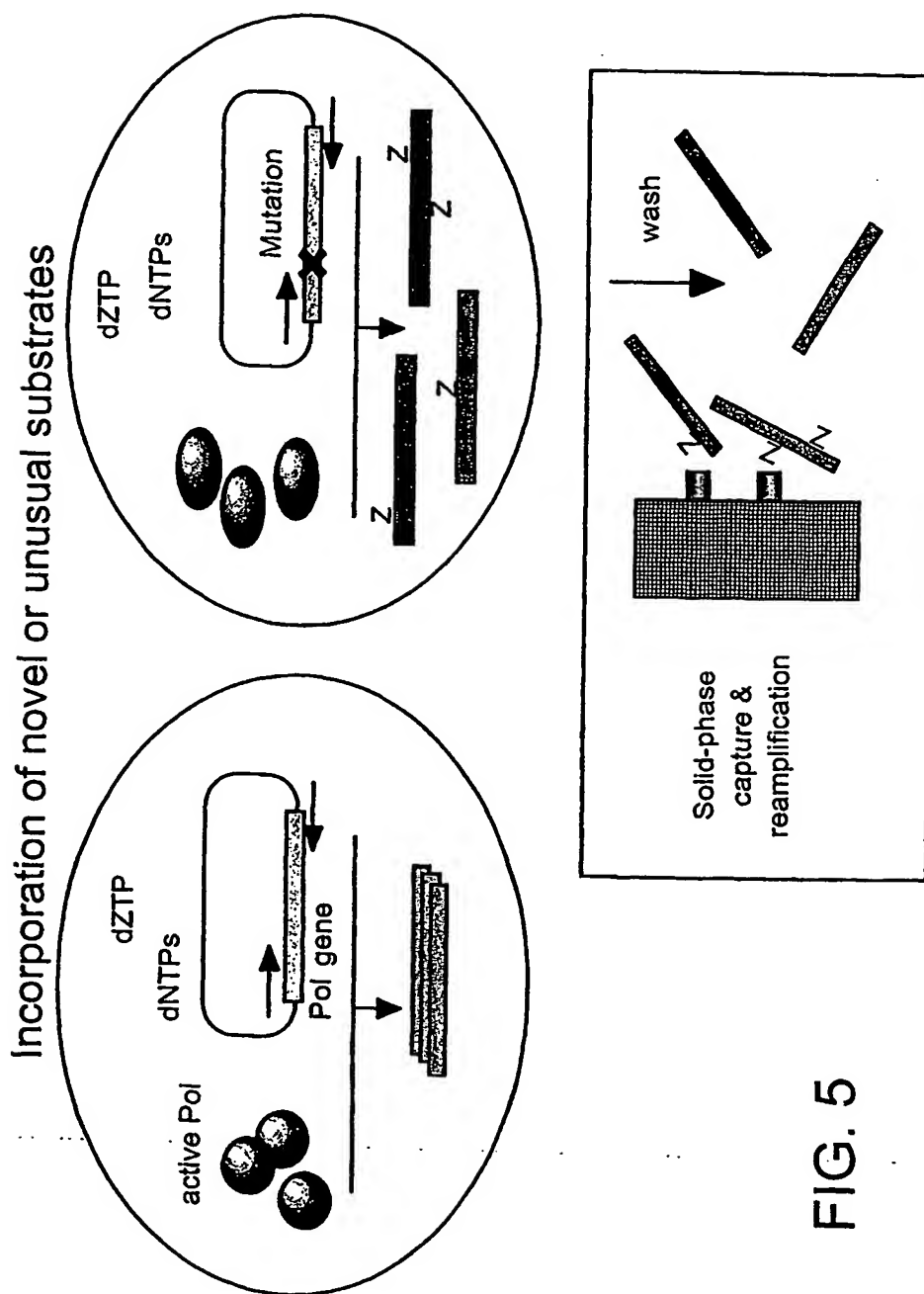


FIG. 5

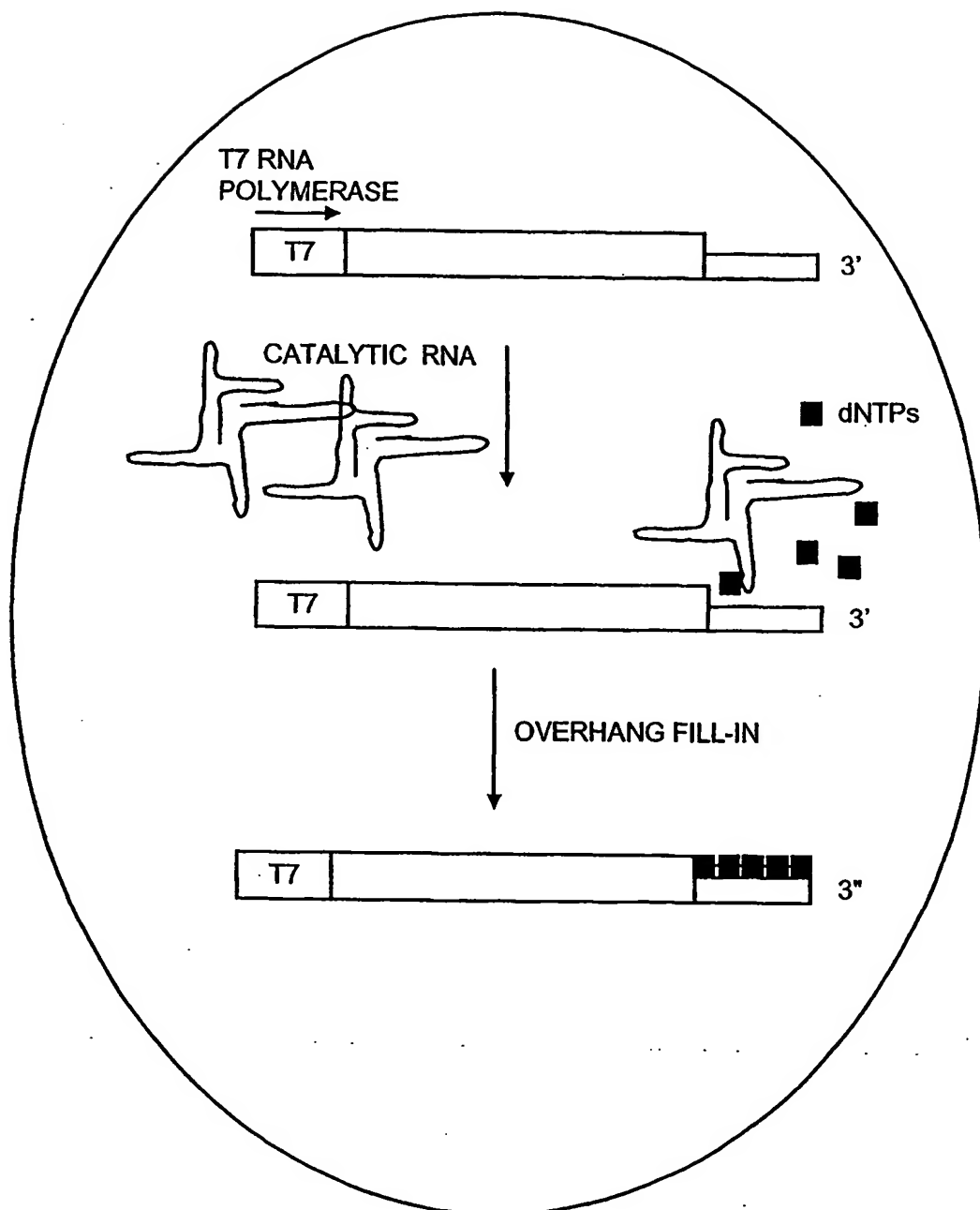


FIG. 6

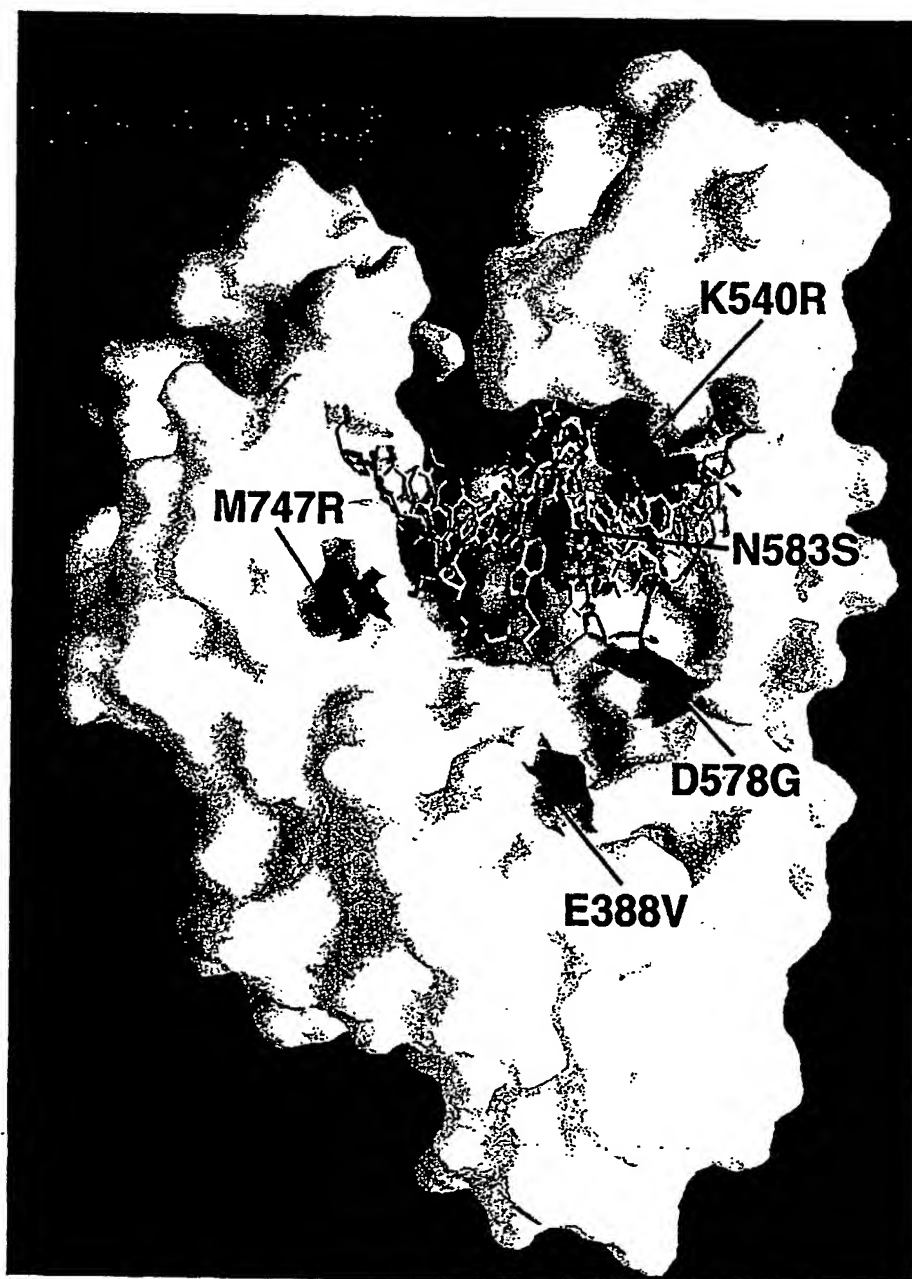


FIG. 7

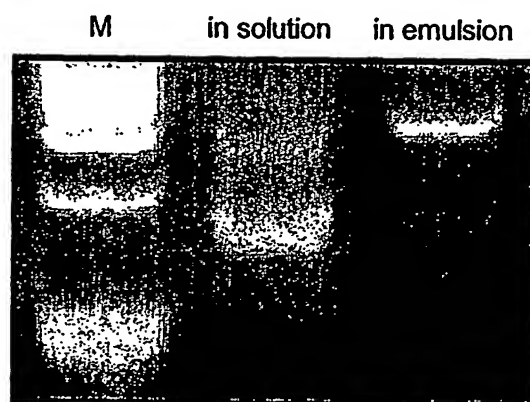
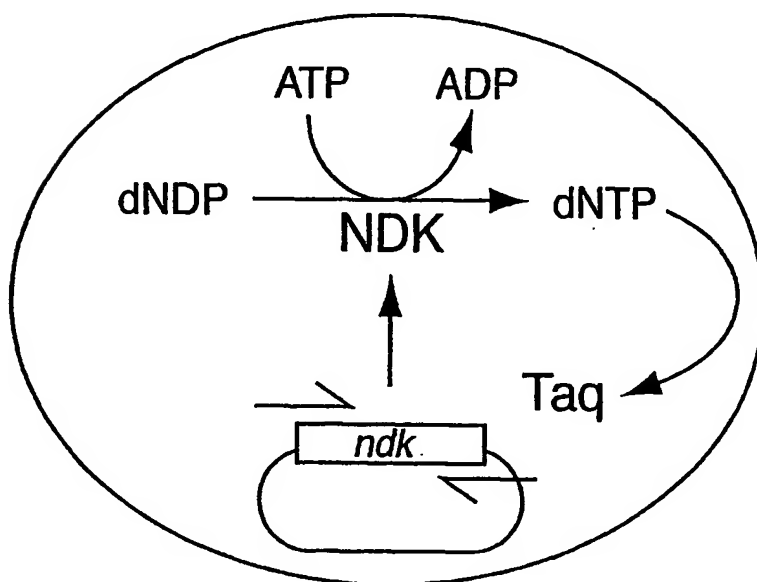


FIG. 8